

Disclosure of Sensitive Behaviors across Self-Administered Survey Modes:

A Meta-Analysis

Timo Gnambs

Osnabrück University

Kai Kaspar

University of Cologne

Author Note

Timo Gnambs, Institute of Psychology, Osnabrück University, Germany. Kai Kaspar, Department of Psychology, University of Cologne, Germany.

We are grateful to Robert Klimanek and Jennifer Lindzus for their aid during the coding process. The Youth Risk Behavior Survey was conducted by the Center for Disease Control and Prevention, the Monitoring the Future study was supported by a grant from the National Institute on Drug Abuse (DA01411), and the National Survey on Drug Use and Health was supported by the Center for Behavioral Health Statistics and Quality (283-2004-00022). Neither of the study sponsors had any role in the study design, analysis, interpretation, or writing of this paper.

Correspondence concerning this article should be addressed to Timo Gnambs, Institute of Psychology, Osnabrück University, Seminarstrasse 20, 49069 Osnabrück, Germany. Email: timo.gnambs@uni-osnabrueck.de

Abstract

In surveys, individuals tend to misreport behaviors being in contrast to prevalent social norms or regulations. Several design features of the survey procedure have been suggested to counteract this problem; particularly, computerized surveys are supposed to elicit more truthful responding. This assumption was tested in a meta-analysis of survey experiments reporting 460 effect sizes (total $N = 125,672$). Self-reported prevalence rates of several sensitive behaviors for which motivated misreporting has been frequently observed were compared across self-administered paper-and-pencil versus computerized surveys. The results revealed that computerized surveys led to significantly more reporting of socially undesirable behaviors than comparable surveys administered on paper. This effect was strongest for highly sensitive behaviors and surveys administered individually to respondents. Moderator analyses did not identify interviewer effects or benefits of audio-enhanced computer surveys. The meta-analysis highlighted the advantages of computerized survey modes for the assessment of sensitive topics.

Keywords: sensitive question, self-disclosure, survey, computer, paper-and-pencil

Disclosure of Sensitive Behaviors across Self-Administered Survey Modes:

A Meta-Analysis

Despite the prominence of self-reports in many areas of social science research, self-reports are prone to various distortions (cf. Chan, 2009), particularly for the assessment of socially undesirable topics such as stigmatized behaviors or illegal activities. Individuals frequently underreport behaviors that are in contrast to prevalent social norms and regulations, even when interviewed in anonymous surveys where respondents do not have to fear negative consequences. For example, typical self-report surveys estimated prevalence rates for smoking that were up to 9 percent points lower than respective rates based on objective biomarkers (Gorber, Schofield-Hurwitz, Hardt, Levasseur, & Tremblay, 2009). To increase the validity of self-reports on sensitive behaviors survey researchers have proposed several solutions (see Tourangeau & Yan, 2007, for a review): among others, the introduction of computerized survey modes has been suggested to increase respondents' anonymity (Buchanan, 2000; Joinson, 1999; Trau, Härtel, & Härtel, 2013) and, as a consequence, should result in more truthful responding. This assumption was examined in a meta-analysis of mode experiments across self-administered paper-and-pencil and computerized surveys for several behaviors conventionally viewed as socially undesirable (e.g., illegal drug use). Moreover, several procedural characteristics associated with the survey process were examined to identify conditions under which computerized surveys are particularly effective in increasing self-disclosure.

Self-Disclosure of Sensitive Behaviors

Sensitive questions address highly personal and sometimes even distressing topics which are often in conflict with social norms and frequently result in socially desirable answers or even non-response. Three aspects can make a question sensitive (Tourangeau, Rips, & Rasinski, 2000): First, a question can be seen as intrusive when it addresses a taboo topic, independent of what the respondent's answers might actually be. Second, fears that

answers to a question might be disclosed to a third party can make it sensitive; particularly if there are concerns about potentially negative consequences associated with a response. Third, questions evoking answers that are in conflict with the prevalent social norm can be perceived as sensitive. Prototypical examples for sensitive topics in many Western societies are the consumption of alcohol and illicit substances (Tourangeau & Yan, 2007), sexual activities (Langhaug, Sherr, & Cowan, 2010; McCallum & Peterson, 2012), or delinquency (Kleck & Roberts, 2012). Due to their private nature researchers interested in studying these behaviors usually have to rely on individuals' self-reports; objective measurements are typically rare (see Pol et al., 2013, for an example on drug use) or nearly impossible (e.g., in the context of sexual research). However, frequently people are reluctant to answer questions they consider sensitive. Even if they relinquish information the validity of their responses is sometimes in question. Data quality does not only depend on the accurate recall of facts but also depends on the degree of peoples' self-disclosure, that is the amount of personal information an individual is willing to provide to others, for example to an interviewer (Jourard, 1971). Self-disclosure is commonly threatened by an individual's inherent need to create and maintain favorable impressions of oneself in the eyes of others (Paulhus, 2002) or, occasionally, to show factitious disorders to excite compassion or interest (Maldonado, 2002). Therefore, respondents tend to misrepresent their true attitudes and behaviors if they believe them to be in conflict with prevalent social norms.

Survey Mode Effects on Self-Disclosure

For a long time, survey researchers have scrutinized factors that might increase self-disclosure of sensitive behaviors (for qualitative reviews see Kleck & Roberts, 2012; Langhaug et al., 2010; McCallum & Peterson, 2012; for quantitative reviews see Richman et al., 1999; Tourangeau & Yan, 2007; Ye, Fulton, & Tourangeau, 2011). Among the studied features, the survey mode was identified as a key variable. A bulk of studies demonstrated that motivated misrepresentation tends to decline for more anonymous surveys that limit

personal interactions with an interviewer (e.g., in telephone surveys) or remove the interviewer entirely from the survey process (e.g., postal surveys). Moreover, computer-administered self-interviews have been suggested to produce even greater self-disclosure as compared to self-administered paper-and-pencil questionnaires because they are presumably perceived as more anonymous (Buchanan, 2000; Joinson, 1999; Trau et al., 2013):

Frequently, computerized conduct evokes an experience of being immersed into another, a virtual, world (cf. also the concept of transportation; Gnambs, Appel, Schreiner, Richter, & Isberner, 2014) letting people forget their immediate surrounding and, thus creates an illusion of privacy; responses seemingly “‘disappear’ into the computer” (Weisband & Kiesler, 1996, p. 3). Therefore, computers are frequently perceived as impartial counterparts reducing respondents’ fear of negative evaluations. The more respondents believe that their responses are not currently being observed by others, the more likely they answer candidly on sensitive issues. Indeed, merely *believing* that computerized responses will not be observed by a human interviewer affect responses, not whether they are actually observed (Lucas, Gratch, King, & Morency, 2014).

Several qualitative reviews supported this assertion and highlighted the advantages of computerized surveys on studies of sexual practices (Langhaug et al., 2010) or delinquent behaviors (Kleck & Roberts, 2012). Even two meta-analyses (Richman et al., 1999; Tourangeau & Yan, 2007) identified small (but generally insignificant) advantages of computer-assisted as compared to paper-and-pencil formats. However, conclusions of the latter are not readily transferable to the assessment of behavioral outcomes: Richman and colleagues (1999) did not examine sensitive behaviors but focused on the social desirability of personality traits, whereas the analyses by Tourangeau and Yan (2007) were based on a rather limited database of only ten samples combining attitudinal, personality, and behavioral scales. Yet another impetus for research on survey mode effects was received with the advent of web-based testing, a variant of computerized surveys administered over the Internet.

According to the “candor” hypothesis (Buchanan, 2000) web-based surveys were assumed to elicit higher self-disclosure because they are perceived as more anonymous. However, existing empirical support for this assumption is inconclusive. Some studies identified the hypothesized effect (e.g., Kays, Gathercoal, & Burow, 2012; Wang et al., 2005), whereas other did not (e.g., Lucia, Herrmann, & Killias, 2007; McCabe, Boyd, Young, Crawford, & Pope, 2005). Thus, hidden moderators might determine the effectiveness of computerized surveys for the disclosure of sensitive information.

Potential Moderators of Mode Effects

Computerized surveys come in many forms (see Couper, 2011, for an overview). For example, some surveys extended traditional computer-assisted formats to audio-enhanced variants in which questions and response options are presented on the computer screen while respondents listen to spoken recordings of the presented item over a headset. Similarly, web-based testing represents a form of unproctored computerized surveying (Gnambs, Batinic, & Hertel, 2011) characterized by specific procedural features (e.g., no direct interaction with an interviewer and no standardized survey setting). Previous research (cf. Aquilino, Wright, & Supple, 2000; Brener et al., 2006; Richman et al., 1999; Tourangeau & Yan, 2007) indicated that a set of survey mode specific conditions associated with the different forms of computerized surveys could moderate the disclosure of sensitive behaviors across survey modes. In addition, mode effects might also depend on specifics of the item content and individual differences of the respondents. Therefore, we examined three groups of moderators referring to item, procedural, or sample characteristics:

Item Sensitivity

Survey respondents are frequently reluctant to discuss sensitive issues with others, particularly people they do not know well (e.g., an interviewer), and refuse to provide answers that might invade their privacy or may violate social norms. As a consequence, response rates to personal questions tend to decrease as the level of sensitivity increases (Bosnjak & Tuten,

2001; Krumpal, 2013; Shoemaker, Eichholz, & Skeews, 2002). Issue sensitivity might also interact with characteristics of the survey process because self-disclosure is strongly connected to the perceived anonymity of the assessment procedure (Joinson, 1999; Joinson, Reips, Buchanan, & Schofield, 2010; Stiglbauer, Gnambs, & Gamsjäger, 2011).

Computerized, particularly web-based, surveys are frequently considered more anonymous and presumably increase the feelings of privacy for the respondents than personal interviews or paper-and-pencil surveys. As a consequence, they yield higher self-disclosure on sensitive topics (Booth-Kewley et al., 2007; Kays et al., 2012). Thus, stronger survey mode differences are expected for the disclosure of highly sensitive behaviors because underreporting of moderately sensitive issues is generally less severe.

Procedural Characteristics

Interviewer presence. Survey mode experiments repeatedly showed that an elimination of the interviewer from the survey process increases self-disclosure of sensitive behaviors (e.g., Chang & Krosnick, 2009, 2010; Ye et al., 2011). Accordingly, Tourangeau and Yan (2007) estimated a median increase of self-reported illicit drug use across seven studies by a factor of 1.3 when the survey was self- as compared to interviewer-administered. It might be speculated that similar effects also manifest in self-administered surveys: the presence of an interviewer might inhibit self-disclosure to some degree if respondents fear that their answers might be accidentally divulged to someone standing nearby. Indeed, there is evidence (Richman et al., 1999) that social desirability effects tend to reduce when respondents are completely alone during test taking (i.e. when no interviewer is present and test taking is conducted alone instead of in group settings). Thus, survey mode differences on self-disclosure are expected to be higher when no interviewer is present during test taking.

Group administration. Bystander effects might contribute to underreporting of sensitive behaviors (Aquilino et al., 2000). If significant others (e.g., parents or spouses) are present during an interview which might be suspected to notice the recorded responses,

underreporting is more likely. For example, experimental studies showed that adolescents underreport their alcohol consumption and marijuana use when their parents were present during the interview (cf. the meta-analysis in Tourangeau and Yan, 2007). Moreover, this effect was qualified by an interaction with the survey mode (cf. Aquilino et al. 2000). The bystander effect was observed in paper-and-pencil surveys whereas computerized forms showed no effect (presumably, because the computer form was perceived as more anonymous). Moreover, the mere presence of others, even if they do not directly interact with a respondent, unconsciously activates goals and perceived norms associated with these individuals (Parks-Stamm, Oettingen, & Gollwitzer, 2010). As a consequence, responses are more likely to reflect prevalent social norms when assessed in group settings. Therefore, surveys administered individually without other test takers being present should result in larger mode differences on the disclosure of sensitive topics than comparable group-administered surveys.

Standardization of setting. Standardized settings create comparable, highly controlled conditions for all respondents, for example by testing in a dedicated lab or room at school. Some authors suggested that standardized survey settings should yield higher prevalence estimates than unstandardized settings testing in respondents' homes (Brener et al., 2006). Fendrich and Johnson (2001) observed in three national surveys on drug abuse that the two surveys being conducted at school resulted in significantly higher prevalence rates of the same behavior than a household survey. This effect was also replicated in respective mode experiments (e.g., Brener et al., 2006; Gfroerer, Wright, & Kopstein, 1997). Adolescents' self-reports of sensitive behaviors were found to result in significantly lower prevalence rates when conducted at home as compared to school settings. However, the pattern of effects is not without dispute because some contradictory evidence has also been found. For example, the hypothesized effect of standardization did not emerge in an experimental study in which respondents were either interviewed at home or in a neutral setting outside home

(Tourangeau, Rasinski, Jobe, Smith, & Pratt, 1997). Moreover, the putative effect of standardization is also at odds with evidence from web-based assessments. Unstandardized surveys administered over the Internet are supposed to increase the perceived anonymity and, thus, facilitate disclosure of sensitive information (e.g., Booth-Kewley et al., 2007; Kays et al., 2012). However, previous research confounded the effects of standardization in web-based research with effects of interviewer presence. To disentangle both effects, the present study will examine the variables as independent moderators.

Audio-enhancements. In audio-enhanced computerized surveys questions and responses are presented on the computer screen while respondents listen to spoken recordings of the presented item over a headset. Audio-enhancement seems to be especially useful to overcome literacy problems for populations with poor reading ability while maintaining high levels of anonymity comparable to traditional computer-assisted surveys (Turner et al., 1998). Existing evidence on the inclusion of an audio component in computerized surveying is mixed. Some studies that compared audio-enhanced computer surveys to interviewer-administered surveys found higher prevalence rates of sensitive behaviors in computerized interviews (e.g., Des Jarlais et al., 1999; Gorbach et al., 2013; Kelly, Soler-Hampejsek, Mensch, & Hewett, 2013; Turner et al., 1998; Yeganeh et al., 2013). However, these studies confounded effects of audio-enhancement with self-administration. Other experimental work comparing different self-administration modes was less clear. Whereas some studies (e.g., Couper, Tourangeau, & Marvin, 2009; Langhaug, Cheung, Pascoe, Hayes, & Cowan, 2009; Tourangeau & Smith, 1996) identified modest benefits of including audio recordings in computer surveys, others did not (e.g., Couper, Singer, & Tourangeau, 2003; Nass, Robles, Bienenstock, Treinen, & Heenan, 2003). Although experimental research was unable to identify a clear pattern of effects for audio-enhancements, a recent qualitative review on self-reported sexual behaviors (Langhaug et al., 2010) concluded that audio-enhanced computer surveys increased self-reports of sexual activities as compared to other self-administered

survey modes. Thus, these results led us to expect larger mode differences in self-disclosure for audio-enhanced computer surveys as compared to traditional computer-assisted survey formats.

Sample Characteristics

Sex of respondents. Although early research on self-disclosure across different survey modes failed to identify significant gender differences (e.g., Miles & Wesley, 1998), more recent studies suggested that male respondents exhibit increased self-disclosure in computerized assessments (Booth-Kewley et al., 2007; Kays, Gathercoal, & Burow, 2012). These sex differences might be a consequence of differences in computer familiarity that tend to be higher for men. They report using the Internet more often (Joiner, Gavin, Duffield, et al., 2005; Joiner, Gavin, Brosnan, et al., 2012) and engaging in more computer-related activities than women (Epstein, 2012). On the other hand, females report more negative attitudes toward computers and the Internet, less computer-related self-efficacy, and more computer-related anxiety (Appel, 2012; Broos, 2005; Hu, Zhang, Dai, & Zhang, 2012). Therefore, it is expected that the increased familiarity with computerized surveys results in an increased likelihood of self-disclosure on sensitive topics for male respondents.

Age of respondents. Compared to adolescents who frequently place less consideration on privacy-related risks, many adults report being more cautious and not to divulge personal information they consider sensitive (e.g., Earp & Baumer, 2003). For example, teenagers are more inclined to provide personal information to businesses (e.g., for marketing purposes) in exchange for minor incentives, for example free gifts (Walrave & Heirman, 2013). The increase of privacy concerns with increasing age becomes particularly evident on the Internet where children and young adults are less concerned about online privacy (Hoofnagle, King, Li, & Turow, 2010). For example, teenagers share more sensitive information such as sexual preferences or political views on social networking sites such as Facebook (Christofides, Muise, Desmarais, 2009, 2012; Walrave, Vanweesenbeck, & Heirman, 2012). These age-

related differences have been attributed to effects of computer-related insecurities that have been shown to increase with age (Laguna & Babcock, 1997). Older individuals tend to report less experience and a lack of confidence with computers (Hawthorn, 2007; Marquie, Jourdan-Boddaert & Huet, 2002). However, this effect seemed to decrease within the last decades (Smith & Oosthuizen, 2006). Thus, it is expected that survey mode effects on self-disclosure are more pronounced for adolescents and young adults than for older age groups.

Present Review

Prevalence rates of sensitive behaviors are examined in a meta-analysis of published mode experiments across paper-and-pencil and computer-assisted survey modes. The meta-analysis complements two related reviews on several important accounts. Whereas Richman and colleagues (1999) primarily studied mode effects with respect to personality and social desirability scales, the present meta-analysis focuses on self-reported behaviors. In addition, new technological advancements made available during the last two decades to survey researchers are taken into account by also including audio-enhanced and web-based surveys, two survey modes that were excluded in Richman et al. (1999). The results in Tourangeau and Yan (2007) are extended by including more than five times as many samples and, more importantly, examining several moderator hypotheses not previously addressed. Thus, the present meta-analysis will provide a more exhaustive understanding of mode effects for computerized surveys than available so far.

The specific hypotheses derived for this meta-analysis are summarized in Table 1. The research focus pertains to computerized survey formats that are expected to yield higher prevalence estimates of self-reported, sensitive behaviors than paper-and-pencil surveys (proposition 1). The difference between survey modes is hypothesized to be contingent on several moderators: survey mode effects are expected to be more pronounced for highly sensitive behaviors (proposition 2) in standardized settings (proposition 3a), when neither an interviewer (proposition 3b) nor other test takers are present during the interview (proposition

3c), and when using computerized surveys including an audio component (proposition 3d).

With regard to characteristics of the respondents, these differences are hypothesized to be most pronounced for adolescent men (propositions 4a and 4b).

Method

Literature Search

Primary studies comparing disclosure of sensitive behaviors in paper-and-pencil and computerized surveys were identified from multiple sources: first, several bibliographic databases (PsycINFO, Psycdex, Psychology & Behavioral Sciences Collection, and EconLit) were searched using the keywords *sensitive questions*, *self-disclosure*, *candor*, *alcohol*, *substance use*, *sexual behavior*, or *delinquency* in combination with *computer-based*, *computerized*, *web-based*, *CASI* or *ACASI*. Second, the respective search was repeated in Google Scholar. Since it seemed infeasible to inspect each of the over 300,000 hits, the search was limited to the first 1,000 results. Because the Google search algorithm ranks search results by importance (Brin & Page, 1998), we are confident to have identified most of the relevant publications from this source. Third, additional studies were taken from the references of previous reviews on social desirability effects in computerized testing (Kleck & Roberts, 2012; Langhaug, Sherr, & Cowan, 2010; McCallum & Peterson, 2012; Richman et al., 1999; Tourangeau & Yan, 2007).

Selection of Sensitive Behaviors

Four rationales guided the selection of sensitive behaviors: first, we focused on socially undesirable practices (e.g., drug use) and did not consider socially desirable behaviors (e.g., voting) because previous research suggested that context factors might differentially affect approach and avoidance behaviors (e.g., Meier, D'Agostino, Elliot, Maier, & Wilkowski, 2012). Second, the behavior should be similarly undesirable across diverse groups of respondents (e.g., being pregnant might be socially undesirable for teenage girls, but seems less undesirable for adult women). Third, because our moderator hypotheses also addressed

potential differences between men and women, sex-specific behaviors (e.g., abortion) were not considered. Finally, we only considered sensitive behaviors that have been routinely examined in previous research (cf. Eaton et al., 2010; Tourangeau & Yan, 2007) and for which relevant effect sizes could be retrieved from published research reports. As a consequence, the meta-analysis focused on four topics conventionally viewed as sensitive (see Table 2): (a) substance use including the consumption of alcohol, tobacco, or illicit drugs (e.g., marijuana, cocaine), (b) sexuality referring to questions about homosexual intercourse, specific sexual practices (e.g., masturbation, oral sex), or sexual activities in exchange for money (e.g., prostitution), (c) delinquency inquiring about carrying a weapon, impersonal offenses (e.g., shoplifting, driving under the influence), or crimes involving physical harm of others (e.g., assault), and (d) victimizations asking about being a victim of physical or sexual abuse, or having attempted suicide.

Inclusion Criteria

A study was included in the meta-analysis when it met the following criteria: (a) the study included a question on at least one of the sensitive behaviors presented in Table 2. (b) The question was administered as self-administered questionnaire in written form on paper and on computer. Studies that compared computerized assessments to personal or telephone interviews were not included. Mode effects for the latter have been reviewed recently by Ye and colleagues (2011; see also De Leeuw & Van der Zouwen, 1988). (c) Participants were either randomly allocated to the two administration modes or provided measures for both modes in a within-subject design. Studies that allowed participants to choose the preferred mode of administration were not included. (d) The assessment procedure was anonymous. Studies that made respondents personally identifiable and linked responses to sensitive questions to specific individuals were excluded. Previous research (e.g., Brown & Venable, 2009; Richman et al., 1999) indicated that mode effects of computerized surveys are limited to anonymous assessment scenarios. (e) Studies on psychiatric patients with severe mental

illness were not considered in order to exclude individuals with impaired cognitive capacity.

(f) The study reported relevant statistics to compute an effect size. This search resulted in 39 primary articles including 48 independent samples (see Table 3).

Moderators

Coded moderators. Several moderators were extracted from the primary studies including four variables that describe features of the assessment procedure (a-d), two sample characteristics (e and f), and the survey year (g): (a) Group administrations were coded as 1 when surveys were administered to groups of test takers (e.g., in a class room). When respondents were alone or respondents could choose their company during the assessment as in web-based testing it was coded as -1. (b) Proctored administrations (coded as 1) where a test administrator supervised the whole testing process and remained present during test taking were contrasted with unproctored administrations (coded as -1) where participants remained alone and unsupervised. (c) Assessment settings that were standardized for all participants (coded as 1)—for example, by testing in a dedicated lab, test center or room at school—were compared to unstandardized settings with varying assessment locations (coded as -1) where each respondents could choose the place to take the survey (e.g., at home or the workplace). (d) The interview type was coded as 1 if the computerized assessment procedure included an audio component and -1 otherwise. Moreover, two sample characteristics that are typically reported in research reports were recorded: (e) the proportion of female participants and (f) the mean age (in years) of the sample. (g) Finally, because the perceived sensitivity of a given topic might change over time (e.g., see Ruel and Campbell, 2006, for the changing stigmatization of HIV), we also extracted the survey year as control variable to examine potential cohort effects. About 29 percent of studies did not report the year of data collection. For these studies the survey year was approximated using the respective publication year. Because the median difference between the survey year and the respective publication year was 3 years for studies reporting both information, the publication year minus 3 was used to

impute missing survey years. The correlations between all moderators are summarized in Table 4.

Sensitivity of behavior. Previous research showed that response rates to personal questions reflect the perceived sensitivity of an item (Bosnjak & Tuten, 2001; Krumpal, 2013; Shoemaker et al., 2002). For example, in an unpublished study by Tourangeau et al. (1997, cited in Tourangeau et al., 2000) demographic items received more valid responses than questions on sexual behaviors. Moreover, non-response to sensitive questions was also a significant predictor of unit non-response, that is, complete study attrition, in panel studies (Loosveldt, Pickery, & Billiet, 2002). Therefore, an objective index reflecting the degree of item sensitivity was derived by examining item non-response in the Youth Risk Behavior Survey (YRBS; Brener et al., 2013), a biannual representative survey ($N \approx 15,000$) on adolescent risk behaviors in the United States. For each sensitive behavior in the YRBS, the percentage of item non-response was estimated. To account for normative differences in behaviors item sensitivity was calculated as the odds ratio of missing responses to the number of affirmative responses. The median of this index from the years 2001 to 2011 was used to guard against potential outliers in a given year. The survey allowed the calculation of sensitivity indices for 15 sensitive behaviors (see Table 2): sensitivity indices were available for substance use and most items on delinquency and victimizations; for sexual behaviors respective indices could not be obtained. The thus calculated index for LSD use fell three standard deviations above the mean index and represented an outlier. Therefore, the presented analyses were limited to the rank information of the sensitivity index. To cross-validate the index we derived a comparable index for 10 behaviors on substance use from the Monitoring the Future studies (MTF; Johnston, Bachman, O'Malley, & Schulenberg, 2011) and the National Surveys on Drug Use and Health (NSDUH; Center for Behavioral Health Statistics and Quality, 2013), annual representative surveys on drug abuse among American youths (MTF; $N \approx 15,000$) or adults (NSDUH; $N \approx 55,000$). The sensitivity rank from the YRBS

correlated with the respective values from the MTF and NSDUH at $r = .94$ and $r = .77^1$. Thus, the derived index showed considerable convergent validity across three independent representative surveys. Consequently, the sensitivity ranks from the YRBS that provided sensitivity information for the largest number of behaviors were used (see Table 2).

Meta-Analytic Procedure

The meta-analysis focused on differences in prevalence rates of risk behaviors; therefore, the odds ratio (OR) was adopted as effect size. The effect sizes were computed as $OR = p_C/p_P$ with p_C as the proportion of respondents agreeing to an item in the computerized survey and p_P as the respective proportion in the paper-and-pencil survey. Therefore, ORs greater than 1 indicated higher prevalence rates and, as such, *higher* self-disclosure in computerized surveys. Using the studentized deleted residual (Viechtbauer & Cheung, 2010), three effects were identified as outliers ($\alpha = .01$), less than 1% of all available ORs. To reduce the impact of these outliers, we followed the approach in Gnambs (2013) and truncated the respective effect sizes to the lower or upper bound of the 90% credibility interval of the true effect calculated from a dataset from which the outliers had been removed.

The effect sizes were aggregated using a random effects meta-analysis (cf. Cheung, 2014a). Following recommendations by Marin-Martínez and Sánchez-Meca (2010), each effect was weighted by the inverse of its variance to account for sampling error. Before calculating these variances, the sample sizes of the ten percent largest studies were truncated to the largest sample size of the remaining studies (cf. Gnambs, 2014). Otherwise, the aggregated effect would primarily reflect the effect of these large-sample studies and give

¹ The somewhat smaller validity correlation for the NSDUH had presumably several reasons. For example, the YRBS and MTF adopted highly standardized assessment settings in dedicated rooms at schools, whereas the NSDUH interviewed respondents at home. Moreover, the two former surveys administered paper-and-pencil questionnaires while the household survey adopted an audio-enhanced computer mode.

hardly any weight to the other studies. Because several studies reported multiple mode comparisons (e.g., obtained for different sensitive behaviors), the meta-analysis was specified as a multilevel model (see Cheung, 2014a). This approach acknowledges the dependencies between the individual effects and models the data on three hierarchical levels: (a) Level 1 refers to the individual effect sizes. (b) Level 2 refers to the effect sizes using different types of sensitive behaviors within a sample; thus, the random level 2 variance $\tau^2_{(2)}$ reflects the heterogeneity of effects due to differences in sensitive behaviors. (c) Level 3 refers to the different samples; thus, the random level 3 variance $\tau^2_{(3)}$ indicates the heterogeneity of effect sizes across samples after controlling for the different types of sensitive behaviors at level 2. The influence of various covariates on the aggregated effect was examined using weighted, mixed-effects regression analyses (Kalaian & Raudenbush, 1996). All analyses were conducted in *R* using the *metaSEM* software (Cheung, 2014b).

Results

Sample Characteristics

The meta-analysis included 48 independent samples (see Table 3) with a total of 125,672 participants (range of the individual studies' *N*s: 27 to 80,515) reporting 460 effect sizes. These samples included, on average, more women than men—the median percentage of female respondents was 59—and primarily comprised of adolescents and young adults, the median age was 19 years. On average, each sample contributed four to five effect sizes. Most effects sizes were available for the comparison of prevalence rates in substance use (65%), whereas the rest focused on victimizations (12%), delinquent (12%) or sexual behaviors (11%). Over two thirds of the studies were conducted in the United States (67%), 15% stem

from Asia, and about 10% originated in European countries². The surveys were administered between the years 1991 and 2010.

Overall Effect of Computerized Assessments

The results of the meta-analysis are summarized in Table 5. The observed, uncorrected odds ratio for all available effect sizes was $OR = 1.24$ which hardly changed after correcting for sampling error, $\Omega = 1.19$. Because the effect sizes were computed in such a way that ORs greater than 1 indicate higher prevalence rates of sensitive behaviors on the computer, these results demonstrated that computerized assessments resulted in significantly, $p < .05$, higher self-disclosure than respective paper-and-pencil modes. This overall effect also replicated for several subgroups of different types of sensitive behaviors. Various forms of substance use, $\Omega = 1.17$, and sexual behaviors, $\Omega = 1.29$, showed significantly, $p < .05$, higher prevalence rates in computerized as compared to paper-and-pencil surveys. Self-reported delinquent behaviors, $\Omega = 1.14$, and victimizations, $\Omega = 1.07$, revealed a similar trend. However, these effects did not reach statistical significance, $p = .09$ and $p = .22$, respectively. Detailed cross-cultural examinations did not seem feasible because very few effects were available from geographical regions outside the United States (see Table 5). But exploratory comparisons of the mean effect sizes calculated for several geographical regions revealed highly similar trends in American, European, African, and Asian samples, with computerized assessments eliciting higher self-disclosure.

Overall, these results support the hypothesized survey mode effect on self-disclosure of sensitive behaviors. However, the significant, $p < .05$, random variances of Ω also pointed at unaccounted heterogeneity that might be accounted for by various moderators.

² The countries (with frequencies in parenthesis) were: Belgium (1), India, (1), Italy (1), Kenya (1), Peru (1), South Africa (1), Switzerland (1), Thailand (2), Taiwan (1), United Kingdom (2), United States (26), and Vietnam (1).

Moderator Analyses

The random variance of the aggregated effect was inspected more closely by meta-regression analysis that used the coded moderators (see method section) as predictors of the individual effect sizes. In these analyses the categorical moderators were contrast (-1 and 1) instead of dummy coded (0 and 1). As a consequence, the intercept in these regression models reflects the mean population effect after controlling for the moderators. Moreover, the continuous moderators (survey year, item sensitivity, sex ratio, and age) were recoded in such a way (as deviations from 2008, 8, 50, and 15, respectively) that the intercept reflects the true mode effect for a behavior of median sensitivity in the year 2008 for samples with a balanced sex ratio and a mean age of 15 years. To guard against potential confounds resulting from cross-cultural differences in self-disclosure (cf. Chen, 1995; Johnson & van de Vijver, 2002) and perceived sensitivity of the studied behaviors (Roster, Albaum, & Smith, 2014) all moderator analyses were limited to the American samples. However, sensitivity analyses including all samples identified highly similar effects.

Survey year. Potential changes across time were examined by modeling the effect sizes dependent on the survey year (see Model 1 in Table 6). Initially, several regression models including higher-order polynomials were also inspected; but only the linear and quadratic terms remained significant, both $p < .06$, and, thus, were retained for the analyses. The effect of computerized assessments on self-disclosure of sensitive behaviors was subject to a moderate time trend (see Figure 1). During the 1990ies mode effects slightly declined and dropped from a predicted $\Omega = 1.25$ to a predicted $\Omega = 1.08$ in the year 2000; the last decade registered a new increase with a predicted $\Omega = 1.19$ in the year 2005. The survey year accounted for about 13 percent of the between-sample heterogeneity $\tau^2_{(3)}$.

Sensitivity of behavior. Sensitivity information was available for a subsample of 283 out of all 343 effects sizes. Regressing these effects on the sensitivity rank, $\gamma = 0.02$, $SE = 0.00$, $p < .01$, highlighted an increase of survey mode differences for more sensitive behaviors

(see Figure 2). This effect was rather robust and remained significant after controlling for the previously identified time trend (see Model 2 in Table 6). Highly sensitive behaviors (predicted $\Omega = 1.63$), such as the use of heroin or cocaine, resulted in larger differences in prevalence rates across survey modes as compared to less sensitive behaviors (predicted $\Omega = 1.43$), such as smoking or the consumption of alcoholic beverages. The sensitivity rank accounted for about 22 percent of the random level 2 variance $\tau^2_{(2)}$. Although the sensitivity of the studied behaviors significantly moderated the survey mode differences, it was not equally predictive for all types of behaviors. For example, as depicted in Figure 2, sexual abuse was classified as a highly sensitive topic. But the empirical, aggregated mode effect was considerably smaller than the predicted effect from the regression model. Thus, additional moderators related to specific types of sensitive behaviors might be unaccounted by the chosen sensitivity index.

Procedural characteristics. Survey mode differences were examined in relation to four procedural characteristics: group administration, interviewer presence, standardization of the survey setting, and inclusion of an audio component. Although some moderators were moderately correlated (see Table 4) variance inflation indices (*VIF*) did not indicate serious multicollinearity (all *VIFs* < 2). Moreover, sensitivity analyses that removed moderators from the regression models one at a time identified the same effects as the full model (Model 3a in Table 6). Among the procedural characteristics, only group administration emerged as a significant moderator; mode differences were more pronounced when respondents were alone without the presence of other test takers (predicted $\Omega = 1.61$) as opposed to settings where other test takers were nearby (predicted $\Omega = 1.18$). Group administration explained $\Delta R^2 = .50$ of the random between-study variance (τ^2_3) in addition to the time trend. The remaining procedural characteristics explained the heterogeneity of effect sizes across studies insufficiently. To examine the robustness of this moderator effect, the respective analyses were also repeated controlling for the item sensitivity. Within the subsample of effects with

sensitivity indices available, the respective moderation effect remained significant, $p < .05$ (see Model 3b in Table 6).

Sample characteristics. For the examination of individual differences between respondents rather few samples were available (about half of all coded samples) because many studies neglected to report relevant socio-demographic information (see Table 3). Moreover, the age range of the available samples was very limited: most studies reported on adolescent samples; in contrast, only two adult samples were available that included respondents with a mean age of 40 years or older. Therefore, the respective analyses should be interpreted with due caution. Moderation analyses (see Model 4 in Table 6) that included the percentage of female participants and the mean age of the studied samples did not identify differences between men and women. However, a marginally significant, $p = .07$, age-related effect emerged. Age explained about $\Delta R^2 = .33$ of the random between-study variance (τ^2_3) in addition to the time trend. Contrary to our expectations samples predominately including adult respondents, predicted $\Omega = 1.45$ at age 30, exhibited stronger self-disclosure in computerized surveys than adolescent samples, predicted $\Omega = 1.29$ at age 15. Because the age of the two adult samples might be considered outliers, we repeated these analyses using the logarithmized age of the respondents as moderator. However, this robustness check failed to replicate the age trend, $p = .12$. Therefore, this result should be regarded as preliminary until a larger body of effects from older respondents is available.

Publication Bias

To determine whether systematically missing studies might have distorted the accuracy of the synthesized effects, Rosenberg's (2005) Fail-Safe N was calculated which indicates the number of studies with null results that one had to add for the estimated Ω to become non-significant. As a rough rule-of-thumb Rosenthal (1979) recommended Fail-Safe N s that are about five times larger than the number of included effects. These indicate robust effects that are unlikely to be distorted by publication bias. As summarized in Table 5, the

estimated Ω for the overall effect can be considered robust. Some authors (e.g., Kepes, Banks, McDaniel, & Whetzel, 2012) evaluated the Fail-Safe N approach for the analysis of publication bias rather critically. Therefore, we also examined the contour-enhanced funnel plot (Peters, Sutton, Jones, Abrams, & Rushton, 2008) including the odds ratios and their standard errors. A visual inspection of the funnel plot (Figure 3) did not indicate publication bias but revealed a largely symmetric distribution around the population effect. Moreover, we also tested the funnel plot statistically for asymmetry by regressing the individual effect sizes on the inverse of their respective sample sizes (cf. Moreno et al., 2009; Peters, Sutton, Jones, Abrams, & Rushton, 2006). A significant effect would indicate funnel plot asymmetry and, thus, a potential publication bias. However, the test failed to identify a significant effect, $B = 7.01$, $SE = 6.43$, $p = .28$ (cf. Table 5), therefore, evidencing no publication bias.

Discussion

Motivated misreporting remains a pervasive problem in survey research, particularly for questions involving behaviors that are contrary to prevalent social norms and, as a consequence, are perceived as embarrassing or even threatening. In these cases, self-reports are more prone to distortions when the specific survey mode requires interpersonal contact with others. Therefore, modes removing the presence of the interviewer from the survey process have been shown to elicit higher self-disclosure of sensitive behaviors than, for example, telephone or personal interviews (cf. Chang & Krosnick, 2009, 2010; Richman et al., 1999; Ye et al., 2011). In addition, it has been suggested that computerization of self-administered surveys would add another level of abstraction leading to even more self-disclosure. Because computers are viewed as impartial communicators that are perceived as more anonymous (e.g., Buchanan, 2000; Joinson, 1999; Richman et al., 1999; Trau et al., 2013), respondents should feel less social pressure to answer in line with prevalent social norms and give more honest answers. In line with this premise, the presented meta-analysis identified significantly higher prevalence rates of sensitive behaviors in computerized as

compared to paper-and-pencil surveys. The respective effect was quite robust and replicated across different types of sensitive behaviors (i.e. substance use, sexuality, delinquency, victimizations) and also different geographical regions. Although the identified mode effect might be considered small, $\Omega = 1.51$ after correcting for several moderators (see Table 6), it was considerably larger than previous research (Tourangeau & Yan, 2007), $\Omega = 1.08$, indicated. However, when point estimates of rare events are of central importance—as in epidemiological research on sensitive topics such as illicit drug use—even the identified small mode effect can be of practical importance, for example when facing costly decisions on the design and implementation of prevention and counseling programs for substance abuse patients.

Interestingly, the studied mode effect showed a marked time trend following an inverted U-shaped function (see Figure 1) that might reflect changes in the respondents' familiarity with the survey technology. Tourangeau and colleagues (2000) suggested the novelty of using computers for interviewing—which was still rather rare in the 1990ies—might have signaled a form of importance and legitimacy for most respondents; in turn, computers might have also increased the disclosure of sensitive behaviors. The increased exposure of respondents to computers might explain the downward trend of this effect in Figure 1. Similarly, the rise of web-based survey modes that gradually gained broader acceptance in psychological research not until the last decade (Gosling, Vazire, Srivastava, & John, 2004) might account for the slight increase in subsequent years.

With regard to the hypothesized moderators (see Table 1), the meta-analysis reached three main conclusions: first, computerization seemed to be particularly advantageous for highly sensitive behaviors such as cocaine use, whereas respective effects were less pronounced for moderately sensitive behaviors, for example smoking or alcohol consumption. Thus, computerized surveying is most effective for the most controversial issues that are strongly in contrast to social norms and regulations. Second, among the studied procedural

survey characteristics co-test takers were most predictive of mode differences. Computerized surveys that were administered alone resulted in significantly higher prevalence estimates of sensitive behaviors than surveys presented to groups of respondents. Thus, traditional web-based surveys seem particularly effective for the collection of sensitive behaviors because test takers can respond alone, without fearing that others might see their responses to sensitive items. Contrary to previous experiments on interracial bias (Evans, Garcia, Garcia, & Baron, 2003), other features of the unproctored computer mode such as the absence of an interviewer did not emerge as an additional moderator. Third, in contrast to some previous findings (e.g., Couper et al., 2009; Langhaug et al., 2009; Tourangeau & Smith, 1996; Turner et al., 1998) computerized surveys experimenting with audio enhancements did not have an additional advantage with regard to self-disclosure. This is somewhat at odds with a recent qualitative review of mode effects in developing countries that reported minor advantages for audio-enhanced computer surveys (Langhaug et al., 2010). The different conclusions from these studies might hint at additional moderators not included in the present meta-analysis. The included moderators accounted for only about half the between-study heterogeneity (see Table 6). Thus, sample characteristics, for example related to the educational level, might explain the discrepant findings. It could be speculated that audio-enhancements would be more effective for specific subgroups with low literacy that were underrepresented in the current meta-analysis.

Overall, the presented results demonstrated that the seemingly minor switch from paper to computer tends to result in higher self-disclosure rates of sensitive behaviors in self-administered surveys.

Accuracy of Self-Reported Sensitive Behaviors

Generally it is assumed that higher prevalence rates of self-reported sensitive behaviors are also more accurate indicators of respondents' real behaviors. However, this "more is better" assumption (Tourangeau & Yan, 2007, p. 863) represents a mostly untested

hypothesis. So far, there are few studies explicitly focusing on the accuracy of self-reported behaviors across survey modes by validating respondents' answers against objective data. The available evidence suggests that the identified increase in prevalence rates is also accompanied by an increase in accuracy (e.g., van Griensven et al., 2006; Hewett et al., 2008; Kreuter, Presser, & Tourangeau, 2008; Langhaug et al., 2010). For example, in a mode experiment Kreuter and colleagues (2008; see also Sakshaug, Yan, & Tourangeau, 2010) validated self-reported academic performance of students against available university records. For socially undesirable questions (e.g., receiving bad grades or having a low grade point average) web-based surveys resulted in significantly less underreporting of true performance than telephone interviews. Similarly, self-reported sexual risk behaviors predicted actual sexually transmitted infections better when respondents were interviewed via audio-enhanced computer surveys as compared to personal interviews (Hewett et al., 2008). Finally, van Griensven and colleagues (2006) validated self-reported substance use including several illicit drugs against objective biomarkers. Descriptive analyses revealed a higher accuracy for computerized assessments than for questionnaires administered on paper. Overall, these studies support the assumption that the different prevalence rates identified for different survey modes are also linked to higher accuracies of these self-reports.

A Matter of Anonymity?

Increased self-disclosure in computerized as compared to paper-and-pencil surveys has been frequently attributed to increases in anonymity perceptions (e.g., Buchanan, 2000; Joinson, 1999; Richman et al., 1999; Trau et al., 2013). However, recent research cast doubts on anonymity as the mediating process because an increase in anonymity can sometimes decrease accountability (Lelkes et al., 2012). Although people tend to report more undesirable behaviors under anonymity conditions the accuracy of the reported behavior decreases. Moreover, many people when given the opportunity to behave unethically also do so (Zhong, Bohns, & Gino, 2010). This is also reflected in the online disinhibition effect resulting in, for

example, a decreased willingness to cooperate with others (Cress & Kimmerle, 2008) or increased flaming behavior (i.e. hostility towards others in web-based communication; Alonzo & Aiken, 2004). Thus, other explanations might account for differences in self-disclosure across self-administered survey modes:

On the one hand, survey mode effects could be a result of increases in confidentiality and privacy (Joinson & Paine, 2006; Joinson et al., 2010). Some survey mode experiments tend to support this notion (DiLillo, DeGue, Kras, DiLoreto-Colgan, & Nash, 2006). Whereas self-administered computerized and paper-and-pencil surveys do not differ with regard to the perceived anonymity, that is whether respondents are personally identifiable and answers to sensitive questions can be linked to specific individuals, the former are perceived as more confidential: computerized modes are attributed with greater privacy, that is whether significant others are expected to see one's responses to sensitive questions. Thus, privacy perceptions, particularly when respondents have control over who gets and does not get access to their responses, seem to increase the willingness to disclose sensitive information (Brandimarte, Acquisti, & Loewenstein, 2012). However, empirical evidence on this point is all but conclusive: it is also conceivable that under certain conditions computerized surveys might be perceived as *less* private, for example when several respondents sitting close to each other might glance at the computer screen of others (Beebe, Harrison, McRae, Anderson, & Fulkerson, 1998; Brener et al., 2006). Moreover, given the ongoing debate on data security and privacy on the Internet future research is highly warranted that scrutinizes the implied mediation mechanism of privacy perceptions on survey modes and self-disclosure.

On the other hand, survey mode effects might be attributed to cognitive distortions in risk perceptions because people tend to underestimate objective risks of events when presented on the computer. For example, many individuals exhibit greater confidence in their abilities (Ackerman & Goldsmith, 2011) and are more likely to hold an illusion of control (i.e. the belief that they can influence even random events; MacKay & Hodgins, 2012) when

identical problems are presented on the computer as compared to other media. Following social-exchange theory (cf. Dillman, Smyth, & Christian, 2014) respondents weight the potential risks in answering a sensitive question against the potential benefits: if the perceived risk outweighs the benefits respondents are more likely to lie or refuse to answer. However, if computerization evokes cognitive distortions that decrease the perceived risk associated with an honest answer, respondents are more likely to disclose a sensitive behavior. As a consequence, prevalence rates of socially undesirable behaviors should be higher in computerized as compared to paper-and-pencil surveys. However, so far, this mediation process has not been examined in the context of survey research and, thus, remains speculative.

Limitations and Outlook

Some limitations might impair the generalization of the presented findings: first, despite showing convergent validity across three large-scale representative surveys, the sensitivity index adopted for this study was not equally capable of predicting survey mode differences for all types of behaviors (e.g., sexual abuse; see Figure 2). Unaccounted confounds might have biased the chosen indicator to some degree. For example, Beatty and Herrmann (2002) argued that item non-response is no pure indicator of item sensitivity. Albeit reflecting the anticipated psychological and social costs of an honest response (i.e. item sensitivity) non-response also reflects respondents' cognitive effort due to item complexity or simply motivational constraints (e.g., a lack of interest). Future research should further scrutinize the domain effect of self-disclosure across survey modes by adopting more elaborate methods, for example, using the randomized response or unmatched count technique (cf. Coutts & Jann, 2011; Lensvelt-Mulders, Hox, Heijden, & Mass, 2005).

Second, respondent characteristics might account for some between-study heterogeneity in the aggregated effect sizes. Socio-demographic characteristics and even personality traits such as an individual's propensity to trust or willingness to take risks could

represent further characteristics differentially affecting reactions to survey computerization. In the present meta-analysis socio-demographic differences were insufficiently able to explain survey mode differences. Although age exhibited a trend-significant effect, this result should be considered with due caution because it is based on rather few samples including predominantly adolescent respondents. Thus, future research should consider systematically examining the sample composition to identify subgroups of respondents for whom computerized survey modes might be particularly effective.

Third, anecdotal evidence also hints at potential mode differences across cultures. For example, North Americans tend to disclose more than Chinese (Chen, 1995), Japanese (Schug, Yuki, & Maddux, 2010) or East Europeans (Maier, Zhang, & Clark, 2013) under face-to-face conditions. However, in computer-mediated environments self-disclosure increases for Asians which has been attributed to the fact that members of collectivistic cultures are more reserved in face-to-face interactions to avoid violating social norms (Zhao, Hinds, & Gao, 2012). Descriptive results could not corroborate these results in the current meta-analysis (see Table 5) because few effects were available from outside the United States. Therefore, future studies are encouraged to explicitly address cultural effects on self-disclosure in computerized surveys.

Finally, the present study was limited to a selection of sensitive behaviors (see Table 2) that has been frequently scrutinized in previous research. We do not want to imply that these are the most important or even only behaviors affected by survey modes. Rather, future research should extend this line research to other content domains that might be considered sensitive such as, for example, political participation (e.g., voting) or self-reported wealth (e.g., income). Indeed, there is evidence that respondents' willingness to report a lower socio-economic status is differentially affected by the survey mode (Pascoe, Hargreaves, Langhaug, Hayes, & Cowan, 2013). Moreover, it might also be worthwhile to extend research on survey mode effects and its moderators to attitudinal questions that dominate public opinion research.

Implications for Survey Research

What are the practical implications of these results? On the one hand, it might be argued that with the widespread availability of web-based and mobile devices (cf. Mavletova & Couper, 2013; Van Heerden, Norris, Tollman, Stein, & Richter, 2014; Wells, Bailey, & Link, 2014) paper-and-pencil surveys will soon become outdated and mode differences should be of no major concern to survey specialists. For example, data from Germany show that in the year 2000 market research firms administered paper-and-pencil surveys about four times more often than computerized formats, whereas this ratio reversed during the subsequent decade; today computerized surveys are administered over four times more often than paper-and-pencil formats (ADM, 2014). Thus, in the near future paper-and-pencil questionnaires might be negligible in survey research. On the other hand, an increasing number of researchers adopt mixed-mode designs which assign respondents to different survey modes to maximize response rates (De Leeuw & Hox, 2011). For example, a study might be designed as a web-based survey; however, to also reach respondents with no or limited Internet access this web-based survey might be supplemented by a postal survey—as, for example, in the nationally representative GESIS panel, a mixed-mode survey of the general population in Germany (cf. Struminskaya, Kaczmirek, Schaurer, & Bandilla, 2014). Given the presented results, the assessment of sensitive behaviors might be biased in mixed-mode surveys when individuals systematically underreport socially undesirable behaviors in paper-and pencil as compared to computer-assisted survey modes.

Conclusions

During the past decades various forms of computerization have been introduced to the survey process, thus, considerably enlarging researchers' degrees of freedom on how to appropriately collect their data (cf. Couper, 2011): from simple paper questionnaires adapted for presentation on computer screens, more sophisticated variants including multimedia components, such as audio or video recordings up to surveys administered over the Internet.

Particularly, web-based surveys have received considerable attention in recent years (e.g., Kays et al., 2012; McCabe et al., 2005), partly because they have been credited with greater anonymity that supposedly should lead to higher self-disclosure of respondents (Buchanan, 2000; Joinson, 1999; Richman et al., 1999; Trau et al.; 2013). The presented meta-analysis seized this assertion and empirically confirmed the effect of survey computerization on the disclosure of sensitive behaviors. Computer-assisted surveys resulted in prevalence rates of sensitive behaviors that were about 1.51 times higher than comparable reports obtained via paper-and-pencil questionnaires; for highly sensitive issues this mode effect was even larger. Thus, surveys on issues conventionally perceived as sensitive tend to benefit from a switch to modern technologies; particularly when respondents are interviewed alone without the presence of other test takers such as in web-based surveys.

References

- Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*, 18-32:
doi:10.1037/a0022086
- Alonzo, M., & Aiken, M. (2004). Flaming in electronic communication. *Decision Support System*, *36*, 205-213. doi:10.1016/S0167-9236(02)00190-2
- *Anastario, M., Chu, H., Soto, E., & Montano, S. (2013). A trial of questionnaire administration modalities for measures of sexual risk behaviour in the uniformed services of Peru. *International Journal of STD & AIDS*, *24*, 513-577.
doi:10.1177/0956462413476273
- Appel, M. (2012). Are heavy users of computer games and social media more computer literate? *Computers & Education*, *59*, 1339-1350. doi:10.1016/j.compedu.2012.06.004
- Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse*, *35*, 845-867. doi:10.3109/10826080009148424
- Arbeitskreis Deutscher Marktforschungsinstitute (ADM). (2014). *Marktforschung in Zahlen 2/2014* [Market research in numbers]. <https://www.adm-ev.de/zahlen/>
- *Bason, J. J. (2000). *Comparison of telephone, mail, web, and IVR surveys of drug and alcohol use among University of Georgia students*. Paper presented at the American Association of Public Opinion Research, Portland, Oregon.
- *Bates, S. C., & Cox, J. M. (2008). The impact of computer versus paper-pencil survey, and individual versus group administration, on self-reports of sensitive behaviors. *Computers in Human Behavior*, *24*, 903-916. doi:10.1016/j.chb.2007.02.021
- Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 71-86). New York, NY: Wiley.

- Beaussart, M. L., & Kaufman, J. C. (2013). Gender differences and the effects of perceived internet privacy on self-reports of sexual behavior and sociosexuality. *Computers in Human Behavior*, *29*, 2524-2529. doi:10.1016/j.chb.2013.06.014
- *Beebe, T. J., Harrison, P. A., McRae, J. A. Jr., Anderson, R. E., & Fulkerson, J. A. (1998). An evaluation of computer-assisted self-interviews in a school settings. *Public Opinion Quarterly*, *62*, 623-632. doi:10.1086/297863
- *Beebe, T. J., Harrison, P. A., Park, E., McRae, J. A., Jr., & Evans, J. (2006). The effects of data collection mode and disclosure on adolescent reporting of health behavior. *Social Science Computer Review*, *25*, 476-488. doi:10.1177/0894439306288690
- *Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior*, *23*, 463-477. doi:10.1016/j.chb.2004.10.020
- Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in web-based surveys. *Journal of Computer-Mediated Communication*, *6*(3). doi:10.1111/j.1083-6101.2001.tb00124.x
- Brandimarte, L., Acquisti, A., & Loewenstein, G. (2012). Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, *4*, 340-347. doi:10.1177/1948550612455931
- *Brener, N. D., Eaton, D. K., Kann, L., Grunbaum, J. A., Gross, L. A., Kyle, T. M., & Ross, J. G. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opinion Quarterly*, *70*, 354-374. doi:10.1093/poq/nfl003
- Brener, N. D., Kann, L., Shanklin, S., Kinchen, S., Eaton, D. K., Hawkins, J., & Flint, K. H. (2013). *Methodology of the Youth Risk Behavior Surveillance System*. Atlanta, GA: Centers for Disease Control and Prevention.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine.

Computer networks and ISDN systems, 30, 107-117. doi:10.1016/S0169-7552(98)00110-

X

Broos, A. M. A. (2005). Gender and information and communication technologies (ICT)

anxiety: Male self-assurance and female hesitation. *CyberPsychology & Behavior*, 8, 21-

31. doi:10.1089/cpb.2005.8.21

*Brown, J. L., & Venable, P. A. (2009). The effects of assessment mode and privacy level on

self-reports of risky sexual behaviors and substance use among young women. *Journal of*

Applied Social Psychology, 39, 2756-2778. doi:10.1111/j.1559-1816.2009.00547.x

Buchanan, T. (2000). Potential of the Internet for personality research. In M. H. Birnbaum

(Ed.), *Psychological experiments on the Internet* (pp. 121-140). San Diego, CA:

Academic Press.

Center for Behavioral Health Statistics and Quality. (2013). *National Survey on Drug Use and*

Health. ICPSR34481-v2. Ann Arbor, MI: Inter-university Consortium for Political and

Social Research [distributor], 2013-06-20. doi:10.3886/ICPSR34481.v2

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance, & R. J.

Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 309-

336). New York, NY: Routledge.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone versus the Internet:

Comparing sample representativeness and response quality. *Public Opinion Quarterly*,

73, 641-78. doi:10.1093/poq/nfp075

Chang, L., & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered

computerized questionnaires: An experiment. *Public Opinion Quarterly*, 74, 154-167.

doi:10.1093/poq/nfp090

Chen, G. (1995). Differences in self-disclosure patterns among Americans vs. Chinese: A comparative study. *Journal of Cross-Cultural Psychology, 26*, 84-91.

doi:10.1177/0022022195261006

Cheung, M. W.-L. (2014a). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*, 211-229.

doi:10.1037/a0032968

Cheung, M. W.-L. (2014b). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods, 46*, 29-40.

doi:10.3758/s13428-013-0361-y

Christofides, E., Muise, A., & Desmarais, S. (2009). Information disclosure and control on Facebook: Are they two sides of the same coin or two different processes?

CyberPsychology & Behavior, 12, 341-345. doi:10.1089/cpb.2008.0226

Christofides, E., Muise, A., & Desmarais, S. (2012). Hey mom, what's on your Facebook? Comparing Facebook disclosure and privacy in adolescents and adults. *Social Psychology and Personality Science, 3*, 48-54. doi:10.1177/1948550611408619

doi:10.1177/1948550611408619

*Chromy, J., Davis, T., Packer, L., & Gfroerer, J. (2002). Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological Issues* (pp. 135-160). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Couper, M. P. (2011). The future of modes of data collection. *Public Opinion Quarterly, 75*, 889-908. doi:10.1093/poq/nfr046

Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of Audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly, 67*, 385-395.

doi:10.1086/376948

Couper, M. P., Tourangeau, R., & Marvin, T. (2009). Taking the audio out of Audio-CASI.

Public Opinion Quarterly, 73, 281-303. doi:10.1093/poq/nfp025

Coutts, E., & Jann, B. (2011). Sensitive questions in online surveys: experimental results for

the Randomized Response Technique (RRT) and the Unmatched Count Technique

(UCT). *Sociological Methods & Research*, 40, 169-193. doi:10.1177/0049124110390768

Cress, U., & Kimmerle, J. (2008). Endowment heterogeneity and identifiability in the

information-exchange dilemma. *Computers in Human Behavior*, 24, 862-874.

doi:10.1016/j.chb.2007.02.022

De Leeuw, E., & Van der Zouwen, J. (1988). Data quality in telephone and face to face

surveys: A comparative meta-analysis. In R. Groves, P. Biemer, L. Lyberg, J. Massey,

W. Nicholls, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New

York, NY: Wiley.

De Leeuw, E. D., & Hox, J. J. (2011). Internet surveys as part of a mixed mode design. In M.

Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet:*

Advances in applied methods and research strategies (pp. 45-76). New York, NY:

Taylor & Francis.

*Denscombe, M. (2006). Web-based questionnaires and the mode effect: An evaluation based

on completion rates and data contents of near-identical questionnaires delivered in

different modes. *Social Science Computer Review*, 24, 245-254.

doi:10.1177/0894439305284522

Des Jarlais, D. C., Paone, D., Milliken, J., Turner, C. F., Miller, H., Gribble, J., ..., &

Friedman, S. R. (1999). Audio-computer interviewing to measure risk behaviour for

HIV among injecting drug users: a quasi-randomised trial. *Lancet*, 353, 1657-1661.

doi:10.1016/S0140-6736(98)07026-3

- *DiLillo, D., DeGue, S., Kras, A., DiLoreto-Colgan, A. R., & Nash, C. (2006). Participant response to retrospective surveys of child maltreatment: Does mode of assessment matter? *Violence & Victims, 21*, 410-424. doi:10.1891/0886-6708.21.4.410
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. New York, NY: Wiley.
- Earp, J. B., & Baumer, D. (2003). Innovative web use to learn about consumer behavior and online privacy. *Communications of the ACM, 46*, 81-83. doi:10.1145/641205.641209
- *Eaton, D. K., Brener, N. D., Kann, L., Denniston, M. M., McManus, T., Kyle, T. M., ..., & Ross, J. G. (2010). Comparison of paper-and-pencil versus web administration of the youth risk behaviour survey (YRBS): Risk behavior prevalence estimates. *Evaluation Review, 34*, 137-153. doi:10.1177/0193841X10362491
- Epstein, J. A. (2012). Factors related to adolescent computer use and electronic game use. *Public Health*, Article ID 795868. doi:10.5402/2012/795868
- Fendrich, M., & Johnson, T. P. (2001). Examining prevalence differences in three national surveys of youth: Impact of consent procedures, mode, and editing rules. *Journal of Drug Issues, 31*, 615-642.
- *Gerbert, B., Bronstone, A., Pantilat, S., McPhee, S., Allerton, M., & Moe, J. (1999). When asked, patients tell: Disclosure of sensitive health-risk behaviors. *Medical Care, 37*, 104-111. doi:10.1097/00005650-199901000-00014
- Gfroerer, J., Wright, D., & Kopstein, A. (1997). Prevalence of youth substance use: the impact of methodological differences between two national surveys. *Drug and Alcohol Dependence, 47*, 19-30. doi:10.1016/S0376-8716(97)00063-X
- Gnambs, T. (2013). The elusive general factor of personality: The acquaintance effect. *European Journal of Personality, 27*, 507-520. doi:10.1002/per.1933

- Gnambs, T. (2014). A meta-analysis of dependability coefficients (test-retest reliabilities) for measures of the Big Five. *Journal of Research in Personality, 52*, 20-28.
doi:10.1016/j.jrp.2014.06.003
- Gnambs, T., Appel, M., Schreiner, C., Richter, T., & Isberner, M.-B. (2014). Experiencing narrative worlds: A latent state-trait analysis. *Personality and Individual Differences, 69*, 187-192. doi:10.1016/j.paid.2014.05.034
- Gnambs, T., Batinic, B., & Hertel, G. (2011). Internetbasierte psychologische Diagnostik [Web-based psychological assessment]. In L. F. Hornke, M. Amelang, & M. Kersting (Eds.), *Verfahren zur Leistungs-, Intelligenz- und Verhaltensdiagnostik*, Enzyklopädie der Psychologie, Psychologische Diagnostik (Vol. II/3, pp. 448-498). Göttingen, Germany: Hogrefe.
- Gorbach, P. M., Mensch, B. S., Husnik, M., Coly, A., Mâsse, B., Makanani, B., ..., & Forsyth, A. (2013). Effect of computer-assisted interviewing on self-reported sexual behavior data in a microbicide clinical trial. *AIDS Behavior, 17*, 790-800.
doi:10.1007/s10461-012-0302-2
- Gorber, S. C., Schofield-Hurwitz, S., Hardt, J., Levasseur, G., & Tremblay, M. (2009). The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine assessed smoking status. *Nicotine & Tobacco Research, 11*, 12-24. doi:10.1093/ntr/ntn010
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93-104. doi:10.1037/0003-066X.59.2.93
- *van Griensven, F., Naorat, S., Kilmarx, P. H., Jeeyapant, S., Manopaiboon, C., Chaikummao, S., ..., & Tappero, J. W. (2006). Palmtop-assisted self-interviewing for the collection of sensitive behavioral data: Randomized trial with drug use urine testing. *American Journal of Epidemiology, 163*, 271-278. doi:10.1093/aje/kwj038

- Hawthorn, D. (2007). Interface design and engagement with older people. *Behaviour and Information Technology*, 26, 333-341. doi:10.1080/01449290601176930
- Hewett, P. C., Mensch, B. S., Ribeiro, M. C. S. D., Jones, H., Lippman, S., Montgomery, M. R., & van de Wijgert, J. (2008). Using sexually transmitted infection biomarkers to validate reporting of sexual behavior within a randomized, experimental evaluation of interviewing methods. *American Journal of Epidemiology*, 168, 202-211. doi:10.1093/aje/kwn113
- Hoofnagle, C., King, J., Li, S., & Turow, J. (2010). *How different are young adults from older adults when it comes to information privacy attitudes and policies?* University of California, Berkeley. doi:10.2139/ssrn.1589864
- Hu, T., Zhang, X., Dai, H., & Zhang, P. (2012). An examination of gender differences among college students in their usage perceptions of the Internet. *Education and Information Technologies*, 17, 315-330. doi:10.1007/s10639-011-9160-1
- *Jaspan, H. B., Flisher, A. J., Myer, L., Mathews, C., Seebregts, C., Berwick, J. R., ..., & Bekker, L.-G. (2007). Methods for collecting sexual behaviour information from South African adolescents - a comparison of paper versus personal digital assistant questionnaires. *Journal of Adolescents*, 30, 353-359. doi:10.1016/j.adolescence.2006.11.002
- *Johnson, A. M., Copas, A. J., Erens, B., Mandalia, S., Fenton, K., Korovessis, C., ..., & Field, J. (2001). Effect of computer-assisted self-interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment. *AIDS*, 15, 111-115. doi:10.1097/00002030-200101050-00016
- Johnson, T., & van de Vijver, F. J. (2002). Social desirability in cross cultural research. In J. Harness, F. J. van de Vijver, & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 193-202). New York, NY: Wiley.

- Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Schulenberg, J. E. (2011). *Monitoring the Future: A continuing study of American youth (12th-Grade Survey)*. ICPSR34409-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-11-20. doi:10.3886/ICPSR34409.v2
- Joiner, R., Gavin, J., Duffield, J., Brosnan, M., Crook, C., Durndell, A., ..., & Lovatt, P. (2005). Gender, Internet identification, and Internet anxiety: Correlates of Internet use. *CyberPsychology & Behavior*, 8, 371-378. doi:10.1089/cpb.2005.8.371
- Joiner, R., Gavin, J., Brosnan, M., Cromby, J., Gregory, H., Guiller, J., ..., & Moon, A. (2012). Gender, Internet experience, Internet identification, and Internet anxiety: A ten-year followup. *Cyberpsychology, Behavior, and Social Networking*, 15, 370-372. doi:10.1089/cyber.2012.0033.
- Joinson, A. N. (1999). Social desirability, anonymity and Internet-based questionnaires. *Behavior Research Methods, Instruments and Computers*, 31, 433-438. doi:10.3758/BF03200723
- Joinson, A. N., & Paine, C. (2006). Self-disclosure, privacy and the Internet. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford Handbook of Internet Psychology* (pp. 237-252). Oxford, United Kingdom: University Press.
- Joinson, A. N., Reips, U. D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. *Human-Computer Interaction*, 25, 1-24. doi:10.1080/07370020903586662
- Jourard, S. M. (1971). *Self-Disclosure: An experimental analysis of the transparent self*. New York, NY: Wiley.
- Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, 1, 227-235. doi:10.1037/1082-989X.1.3.227

- Kays, K., Gathercoal, K., & Burhow, W. (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior, 28*, 251-256.
doi:10.1016/j.chb.2011.09.007
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*, 624-662.
doi:10.1177/1094428112452760
- Kelly, C. A., Soler-Hampejsek, E., Mensch, B. S., & Hewett, P. C. (2013). Social desirability bias in sexual behavior reporting: Evidence from an interview mode experiment in rural Malawi. *International Perspectives on Sexual and Reproductive Health, 39*, 14-21. doi:10.1363/3901413
- Kleck, G., & Roberts, K. (2012). What survey modes are most effective in eliciting self-reports of criminal or delinquent behavior? In L. Gideon (Ed.), *Handbook of Survey Methodology in Social Sciences* (pp. 417-439). New York, NY: Springer.
- *Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behavior, 19*, 117-134. doi:10.1016/S0747-5632(02)00008-0
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys. *Public Opinion Quarterly, 72*, 847-865. doi:10.1093/poq/nfn063
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity, 47*, 2025-2047. doi:10.1007/s11135-011-9640-9
- Laguna, K., & Babcock, R. L. (1997). Computer anxiety in young and older adults: implications for human-computer interactions in older populations. *Computers in Human Behavior, 13*, 3, 317-326. doi:10.1016/S0747-5632(97)00012-5
- Langhaug, L. F., Cheung, Y. B., Pascoe, S., Hayes, R., & Cowan, R. M. (2009). Differences in prevalence of common mental disorder as measured using four questionnaire

- delivery methods among young people in rural Zimbabwe. *Journal of Affective Disorders*, *118*, 220-223. doi:10.1016/j.jad.2009.02.003
- Langhaug, L. F., Sherr, L., & Cowan, F. M. (2010). How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. *Tropical Medicine and International Health*, *15*, 362-381. doi:10.1111/j.1365-3156.2009.02464.x
- *Le, L. C., Blum, R. W., Magnani, R., Hewett, P. C., & Mai, H. (2006). A pilot of audio computer-assisted self-interview for youth reproductive health research in Vietnam. *Journal of Adolescent Health*, *38*, 740-747. doi:10.1016/j.jadohealth.2005.07.008
- Lelkes, Y, Krosnick, J., Max, D., Judd, C., & Park, B. (2012). Complete anonymity compromises the accuracy of self-reports. *Journal of Experimental Social Psychology*, *48*, 1291-1299. doi: 10.1016/j.jesp.2012.07.002
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., Heijden, P. G. M. van der, & Mass, C. J. M. (2005). Meta-analysis of randomized response research. Thirty-five years of validation. *Sociological Methods Research*, *33*, 319-348. doi:10.1177/0049124104268664
- *Link, M. W., & Mokdad, A. H. (2005). Effects of survey mode on self-reports of adult alcohol consumption: A comparison of mail, web and telephone approaches. *Journal of Studies on Alcohol*, *66*, 239-245.
- Loosveldt, G., Pickery, J., & Billiet, J. (2002). Item nonresponse as predictor of unit nonresponse in a panel survey. *Journal of Official Statistics*, *18*, 545-557.
- Lucas, G. M., & Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94-100. doi:10.1016/j.chb.2014.04.043
- *Lucia, S., Herrmann, L., & Killias, M. (2007). How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different

definitions of the reference period. *Journal of Experimental Criminology*, 3, 39-64.

doi:10.1007/s11292-007-9025-1

*Lygidakis, C., Rigon, S., Cambiaso, S., Bottoli, E., Cuzzo, F., Bonetti, S., ..., & Marzo, C.

(2010). A web-based versus paper questionnaire on alcohol and tobacco in adolescents.

Telemedicine and e-Health, 16, 925-930. doi:10.1089/tmj.2010.0062

MacKay, T.-L., & Hodgins, D. C. (2012). Cognitive distortions as a problem gambling risk

factor in Internet gambling. *International Gambling Studies*, 12, 163-175.

doi:10.1080/14459795.2011.648652

Maier, G. A., Zhang, Q., & Clark, A. (2013). Self-disclosure and emotional closeness in

intracultural friendships: A cross-cultural comparison among U.S. Americans and

Romanians. *Journal of Intercultural Communication*, 42, 22-34.

doi:10.1080/17475759.2012.703620

Maldonado, J. R. (2002). When patients deceive doctors: a review of factitious disorders.

American Journal of Forensic Psychiatry, 23, 29-58.

Marquie, J. C., Jourdan-Boddaert, L., & Huet, N. (2002). Do older adults underestimate their

actual computer knowledge? *Behaviour and Information Technology*, 21, 4, 273-280.

doi:10.1080/0144929021000020998

Mavletova, A., & Couper, M.P. (2013). Sensitive topics in PC web and mobile web surveys:

Is there a difference? *Survey Research Methods*, 7, 191-205.

*McCabe, S. E., Boyd, C. J., Couper, M. P., Crawford, S., & D'Arcy, H. (2002). Mode effects

for collecting alcohol and other drug use data: Web and U.S. mail. *Journal of Studies on*

Alcohol, 63, 755-761.

*McCabe, S. E. (2004). Comparison of web and mail surveys in collecting illicit drug use

data: a randomized experiment. *Journal of Drug Education*, 34, 61-72.

doi:10.2190/4hey-vwxl-dvr3-hakv

- *McCabe, S. E., Boyd, C. J., Young, A., Crawford, S., & Pope, D. (2005). Mode effects for collecting alcohol and tobacco data among 3rd and 4th grade students: A randomized pilot study of web-form versus paper-form surveys. *Addictive Behaviors, 30*, 663-671. doi:10.1016/j.addbeh.2004.08.012
- McCallum, E. B., & Peterson, Z. D. (2012). Investigating the impact of inquiry mode on self-reported sexual behavior: Theoretical considerations and review of the literature. *Journal of Sex Research, 49*, 212-226. doi:10.1080/00224499.2012.658923
- Meier, B. P., D'Agostino, P. R., Elliot, A. J., Maier, M. A., & Wilkowski, B. M. (2012). Color in context: Psychological context moderates the influence of red on approach-and avoidance-motivated behavior. *PLoS ONE, 7*, e40333. doi:10.1371/journal.pone.0040333
- *Mensch, B. S., Hewett, P. C., & Erulkar, A. (2003). The reporting of sensitive behavior among adolescents: A methodological experiment in Kenya. *Demography, 40*, 247-268. doi:10.1353/dem.2003.0017
- Miles, E., & Wesley, K. (1998). Gender and administration mode effects when pencil-and-paper personality tests are computerized. *Educational and Psychological Measurement, 58*, 68-76. doi:10.1177/0013164498058001006
- Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology, 9*. doi:10.1186/1471-2288-9-2
- *Morrison-Beety, D., Carey, M. P., & Tu, X. (2006). Accuracy of audio computer-assisted self-interviewing (ACASI) and self-administered questionnaires for the assessment of sexual behavior. *AIDS Behavior, 10*, 541-552. doi:10.1007/s10461-006-9081-y
- Nass, C., Robles, E., Heenan, C., Bienenstock, H., & Treinen, M. (2003). Speech-based disclosure systems: Effects of modality, gender of prompt, and gender of user.

International Journal of Speech Technology, 6, 113-21.

doi:10.1023/A:1022378312670

*Onoye, J. M., Goebert, D. A., & Nishimura, S. T. (2012). Use of incentives and web-based administration for surveying student alcohol and substance use in an ethnically diverse sample. *Journal of Substance Use*, 17, 61-71. doi:10.3109/14659891.2010.526167

*O'Reilly, J. M., Hubbard, M. L., Lessler, J. T., Biemer, P. P., & Turner, C. F. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, 10, 197-214.

Parks-Stamm, E. J., Oettingen, G., & Gollwitzer, P. M. (2010). Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving. *Journal of Experimental Social Psychology*, 46, 531-542. doi:10.1016/j.jesp.2010.02.004

Pascoe, S. J. S., Hargreaves, J. R., Langhaug, L. F., Hayes, R. J., & Cowan, F. M. (2013). 'How poor are you?' - A comparison of four questionnaire delivery modes for assessing socio-economic position in rural Zimbabwe. *PLoS ONE*, 8, e74977.

doi:10.1371/journal.pone.0074977

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaum.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association*, 295, 676-680. doi:10.1001/jama.295.6.676

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61, 991-996.

doi:10.1016/j.jclinepi.2007.11.010

- Pol, P. van der, Liebrechts, N., Graaf, R. de, Korf, D. J., Brink, W. van den, & Laar, M. van den (2013). Validation of self-reported cannabis dose and potency: an ecological study. *Addiction, 108*, 1801-1808. doi:10.1111/add.12226
- *Potdar, R., & Koenig, M. A. (2005). Does Audio-CASI improve reports of risky behavior? Evidence from a randomized field trial among young urban men in India. *Studies in Family Planning, 36*, 107-116. doi:10.1111/j.1728-4465.2005.00048.x
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interview. *Journal of Applied Psychology, 84*, 754-775. doi:10.1037/0021-9010.84.5.754
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*, 464-468. doi:10.1554/04-602
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641. doi:10.1037/0033-2909.86.3.638
- Roster, C. A., Albaum, G., & Smith, S. M. (2014). Topic sensitivity and Internet survey design: A cross-cultural/national study. *Journal of Marketing Theory and Practice, 22*, 91-102. doi:10.2753/MTP1069-6679220106
- Ruel, E., & Campbell, R.T. (2006). Homophobia and HIV/AIDS: Attitude change in the face of an epidemic. *Social Forces, 84*, 2167-2178. doi:10.1353/sof.2006.0110
- *Rumakom P., Guest, P., Chinvarasopak, W., Utarmat, W., & Sonta-nakanit, J. (2005). Obtaining accurate responses to sensitive questions among Thai students: a comparison of two data collection techniques. In S. Jejeebhoy, I. Shah, & S. Thapa (Eds.), *Sex Without Consent* (pp. 318-332). London, United Kingdom: Zed Books.

Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: Tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly*, *74*, 907-933. doi:10.1093/poq/nfq057

*SAMHSA (2001). *Development of computer-assisted interviewing procedures for the National Household Survey on Drug Abuse*. Substance Abuse and Mental Health Services Administration (SAMHSA), Department of Health and Human Services, Rockville, MD.

*Sarrazin, M. S. V., Hall, J. A., Richards, C., & Carswell, C. (2002). A comparison of computer-based versus pencil-and-paper assessment of drug use. *Research on Social Work Practice*, *12*, 669-683. doi:10.1177/1049731502012005006

Schug, J., Yuki, M., & Maddux, W. (2010). Relational mobility explains between- and within-culture differences in self-disclosure to close friends. *Psychological Science*, *21*, 1471-1478. doi:10.1177/0956797610382786

Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, *14*, 193-201. doi:10.1093/ijpor/14.2.193

Smith, E., & Oosthuizen, H. J. (2006). Attitudes of entry-level university students towards computers: A comparative study. *Computers and Education*, *47*, 352-371. doi:10.1016/j.compedu.2004.10.011

Stiglbauer, B., Gnambs, T., & Gamsjäger, M. (2011). The interactive effects of motivations and trust in anonymity on adolescents' enduring participation in web-based social science research: A longitudinal behavioral analysis. *International Journal of Internet Science*, *6*, 29-43.

Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. (2014). Assessing representativeness of a probability-based online panel in Germany. In M. Callegaro, R.

- Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 62-85). West Sussex, England: Wiley.
- *Supple, A. J., Aquilino, W. S., & Wright, D. L. (1999). Collecting sensitive self-report data with laptop computers: Impact on the response tendencies of adolescents in a home interview. *Journal of Research on Adolescence*, 9, 467-488.
doi:10.1207/s15327795jra0904_5
- *Testa, M., Livingston, J. A., & VanZile-Timsen, C. (2005). The impact of questionnaire administration mode on response rate and reporting of consensual and nonconsensual sexual behavior. *Psychology of Women Quarterly*, 29, 345-352. doi:10.1111/j.1471-6402.2005.00234.x
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, United Kingdom: Cambridge University Press.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60, 275-304. doi:10.1086/297751
- Tourangeau, R., Rasinski, K., Jobe, J., Smith, T. W., & Pratt, W. (1997). Sources of error in a survey of sexual behavior. *Journal of Official Statistics*, 13, 341-365.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859-883. doi:10.1037/0033-2909.133.5.859
- Trau, R. N. C., Härtel, C. E. J., & Härtel, G. F. (2013). Reaching and hearing the invisible: Organizational research on invisible stigmatized groups via web surveys. *British Journal of Management*, 24, 532-541. doi:10.1111/j.1467-8551.2012.00826.x
- *Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, 280, 867-873.
doi:10.1126/science.280.5365.867

- Van Heerden, A. C., Norris, S. A., Tollman, S. M., Stein, A. D., & Richter, L. M. (2014). Field lessons from the delivery of questionnaires to young adults using mobile phones. *Social Science Computer Review*, 32, 105-112. doi:10.1177/0894439313504537
- *Vereecken, C. A., & Maes, L. (2006). Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. *Journal of Adolescent Health*, 38, 426-432. doi:10.1016/j.jadohealth.2004.10.010
- Viechtbauer, W., & Cheung, W. (2010). Outlier and influencer diagnostics for meta-analysis. *Research Synthesis Methods*, 1, 110-125. doi:10.1002/jrsm.11
- Walrave, M., & Heirman, W. (2013). Adolescents, online marketing and privacy: predicting adolescents' willingness to disclose personal information for marketing purposes. *Children & Society*, 27, 434-447. doi:10.1111/j.1099-0860.2011.00423.x
- Walrave, M., Vanwesenbeeck, I., & Heirman, W. (2012). Connecting and protecting? Comparing predictors of self-disclosure and privacy settings use between adolescents and adults. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 6, article 3. doi:10.5817/CP2012-1-3
- *Wang, Y.-C., Lee, C.-M., Lew-Ting, C.-Y., Hsiao, C. K., Chen, D. R., & Chen, W. J. (2005). Survey of substance use among high school students in Taipei: Web-based questionnaire versus paper-and-pencil questionnaire. *Journal of Adolescent Health*, 37, 289-295. doi:10.1016/j.jadohealth.2005.03.017
- Weisband, S., & Kiesler, S. (1996). Self-disclosure on computer forms: Meta-analysis and implications. R. Bilger, S. Guest, & M. J. Tauber (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3-10). Vancouver, CA: ACM. doi:10.1145/238386.238387
- Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review*, 32, 238-255. doi:10.1177/0894439313505829

- *Wright, D. L., Aquilino, W., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol and drug use. *Public Opinion Quarterly*, *62*, 331-353. doi:10.1086/297849
- *Wu, Y., & Newfield, S. A. (2007). Comparing data collected by computerized and written surveys for adolescence health research. *Journal of School Health*, *77*, 23-28. doi:10.1111/j.1746-1561.2007.00158.x
- Ye, C., Fullton, J., & Tourangeau, R. (2011). More positive or more extreme? A meta-analysis of mode differences in response choice. *Public Opinion Quarterly*, *75*, 349-365. doi:10.1093/poq/nfr009
- Yeganeh, N., Dillavou, C., Simon, M., Gorbach, P., Santos, B., Fonseca, R., ..., & Nielsen-Saines, K. (2013). Audio computer-assisted survey instrument versus face-to-face interviews: optimal method for detecting high-risk behaviour in pregnant women and their sexual partners in the south of Brazil. *International Journal of STD & AIDS*. Advance online publication. doi:10.1177/0956462412472814
- Zhao, C., Hinds, P., & Gao, G. (2012). How and to whom people share: The role of culture in self-disclosure in online communities. In S. Poltrock, & C. Simone (Eds.), *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 67-76). New York, NY: ACM. doi:10.1145/2145204.2145219
- Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increase dishonesty and self-interested behavior. *Psychological Science*, *21*, 311-314. doi:10.1177/0956797609360754

* Article included in the meta-analysis.

Table 1

Overview of Study Propositions

Proposition	
	Surveys yield higher prevalence rates of sensitive behaviors ...
1.	... when administered on computer than on paper
	Differences in prevalence rates of sensitive behaviors are larger ...
	<i>Item sensitivity</i>
2.	... for highly sensitive as compared to moderately sensitive behaviors
	<i>Procedural characteristics</i>
3a.	... when surveys are administered in standardized settings
3b.	... when no interviewer is present during survey administration
3c.	... when surveys are administered alone without the presence of other test takers
3d.	... for computerized surveys incorporating audio-enhancements
	<i>Sample characteristics</i>
4a.	... for predominantly male samples
4b.	... for samples with predominantly younger individuals

Table 2

Examples of Sensitive Questions with Sensitivity Indices

			Sensitivity	
<i>Substance use</i>			Index	Rank
1.	AL	Alcohol (e.g., beer or wine)	0.13	5
2.	TO	Cigarettes or cigars	0.07	3
3.	MA	Marijuana	0.06	2
4.	CO	Cocaine or crack	0.19	10
5.	IN	Inhalants (e.g., sniffed glue)	0.16	8
6.	HE	Heroin	0.63	14
7.	ME	Methamphetamines (speed)	0.14	6
8.	EC	Ecstasy	0.30	11
9.	LS	LSD	1.49	15
10.	MM	Misuse of medicaments (e.g., sedatives, tranquilizers)	0.36	12
<i>Sexuality</i>				
11.	HI	Homosexual intercourse		
12.	SE	Specific sexual practices (e.g., oral sex)		
13.	BO	Bought or sold sex		
<i>Delinquency</i>				
14.	WE	Carried a weapon (e.g., gun or knife)	0.15	7
15.	DR	Drove a car under the influence	0.18	9
16.	IM	Impersonal offenses (e.g., shoplifting)		
17.	FI	Personal offenses (e.g., fighting)	0.05	1
<i>Victimization</i>				
18.	PA	Physical abuse		
19.	SA	Sexual abuse (e.g., forced to have sex)	0.43	13
20.	SU	Suicide plan or attempt	0.09	4

Note. The sensitivity index was calculated as the ratio of item non-response to the number of affirmative responses in the Youth Risk Behavior Survey (Brener et al., 2013). The median of this index from the years 2001 to 2011 is reported. The index for LSD use represents an outlier (i.e. falling three *SD* above *M*). Higher indices and ranks indicate more sensitive questions.

Table 3

Summary of Samples included in the Meta-Analysis

Source	Survey year	Sample characteristics			Procedural characteristics					
		Country	♀	Age	G	I	S	A	<i>N</i>	<i>k</i>
Astario et al. (2013)	2010	Peru	22	31	i	p	s	i	332	2
Bason (2000)	2000	US	66	22	g	n	u	n	319	7
Bates & Cox (2008)		US	62		i	p	s	n	50	4
		US	62		g		s	n	43	4
		US	62		g	n	u	n	44	4
Beebe et al. (1998)	1996	US			i	p	s	n	368	15
Beebe et al. (2006)	2000	US	52	15	g	n	s	n	408	11
Booth-Kewley et al. (2007)		US	0	19	i	p	s	n	108	1
		US	100	19	i	p	s	n	193	1
Brener et al. (2006)	2004	US	58	16	i	p	s	n	2,297	25
	2004	US	55	16	g	p	u	n	2,209	25
Brown & Venable (2009)		US	100	20	g	p	s	n	100	5
Chromy et al. (2002)	1999	US	69		g	p	u	i	80,515	28
Denscombe (2006)	2004	England			i	p	s	n	338	1
DiLillo et al. (2006)		US	100	20	g	p	s	n	226	2
Eaton et al. (2010)	2008	US	51	15	i	p	s	n	5,227	39
Gerbert et al. (1999)		US	62	40	g	p	s	n	780	6
van Griensven et al. (2006)	2002	Thailand	50		g	p	s	n	271	28
Jaspan et al. (2007)		South Africa	68	15	g	p	s	n	166	3
Johnson et al. (2001)	1998	England	59		g	p	u	n	829	5
Knapp & Kirk (2003)	1999	US	78	22	g	n	u	n	231	7
Le et al. (2006)		Vietnam	0	20	g	p	u	i	739	1
Link & Mokdad (2005)		US	64	50	g	n	u	n	1,979	2
Lucia et al. (2007)	2004	Switzerland			i	p	s	n	1,203	33
Lygidakis et al. (2010)		Italy	0	15	i	p	s	n	96	4
		Italy	100	15	i	p	s	n	94	4
McCabe (2004)	2001	US	100		g	n	u	n	2,109	20
	2001	US	0		g	n	u	n	1,497	20
McCabe et al. (2005)	2003	US	45		i	p	s	n	280	2
Mensch et al. (2001)	2000	Kenya	0	18	g	p	u	i	1,444	2
	2000	Kenya	100	18	g	p	u	i	1,361	2

Table 3 (continued).

Source	Survey year	Sample characteristics			Procedural characteristics					
		Country	♀	Age	G	I	S	A	<i>N</i>	<i>k</i>
Morrison-Beedy et al. (2006)	2003	US	100	20	g	p	s	i	51	4
	2003	US	100	20	g	p	s	i	51	4
Onoye et al. (2012)	2006	US	56		g	p	s	n	1,531	7
O'Reilly et al. (1994)		US	86		g	p	s	i	27	15
Potdar & König (2005)	2003	India	0	19	i	p	s	i	600	12
Rumakom et al. (2005)		Thailand	0	21	i	p	s	i	197	4
		Thailand	100	20	i	p	s	i	249	4
SAMHSA (2001)	1997	US	57		g	p	s	i	5,070	27
Sarrazin et al. (2002)	1996	US			g	p	s	n	99	2
Supple et al. (1999)	1995	US	51	15	g	p	u	n	1,072	10
Testa et al. (2005)	2002	US	100	24	g			n	1,332	3
Turner et al. (1998)	1995	US	0	17	g	p	u	i	1,711	19
Veerecken & Maes (2004)	2000	Belgium	0	15	i	p	s	n	900	2
	2000	Belgium	100	15	i	p	s	n	708	2
Wang (2005)	2003	Taiwan	38		i	p	s	n	1,918	8
Wright et al. (1998)	1995	US	54		i	p	s	n	3,169	12
Wu & Newfield (2007)	2002	US	58	15	g	p	u	i	1,131	12

Note. Sample characteristics: ♀ = Percentage female, Age = Mean age in years; Procedural characteristics: G = Group administration (g = group, i = individual), I = Interviewer presence (p = present, n = not present), S = Survey setting (s = standardized, u = unstandardized), A = Audio-enhancement (i = included, n = not included); *k* = Number of effects.

Table 4

Correlations between Moderators

	<i>Mdn / %</i>	1.	2.	3.	4.	5.	6.	7.
1. Survey year	2002							
2. Sensitivity of behavior	5	-.02						
3. Group administration 1 = group -1 = individual	26% 74%	.20	-.16					
4. Interviewer presence 1 = present -1 = not present	76% 24%	-.05	.10	.35				
5. Survey setting 1 = standardized -1 = unstandardized	63% 37%	.12	.03	.46*	.56*			
6. Audio-enhancement 1 = available -1 = not available	23% 77%	-.35	.11	-.33	.33	-.06		
7. Percentage of female respondents	62	.13	.23	-.21	.04	.22	.06	
8. Mean age of respondents	20	-.19	.15	-.25	-.36	-.11	-.16	.12

Note. $k = 16$ to 31 US samples

* $p < .05$

Table 5

Meta-Analysis of Sensitive Questions in Computerized Assessments

				Observed effect			True effect							
	k_1	k_2	N	OR	logOR	SD	Ω	log Ω	SE	95% CI	$\tau^2_{(2)}$	$\tau^2_{(3)}$	$I^2_{(2)}$	$I^2_{(3)}$
Overall	460	48	125,305	1.24	0.22	0.52	1.19	0.17*	0.04	[0.10, 0.25]	0.03*	0.04*	.29	.38
<i>Type of sensitive behavior</i>														
Substance use	300	38	121,367	1.26	0.23	0.52	1.17	0.15*	0.04	[0.07, 0.23]	0.03*	0.04*	.30	.39
Sexuality	51	20	17,966	1.44	0.37	0.72	1.29	0.26*	0.08	[0.11, 0.41]	0.07 [†]	0.01	.42	.07
Delinquency	53	12	14,410	1.13	0.12	0.34	1.14	0.14 [†]	0.08	[-0.02, 0.30]	0.04*	0.04	.34	.39
Victimization	56	17	15,964	1.09	0.09	0.39	1.07	0.07	0.06	[-0.04, 0.18]	0.00	0.02	.00	.35
<i>Geographical region</i>														
United States	343	31	113,837	1.23	0.21	0.50	1.17	0.15*	0.05	[0.06, 0.24]	0.03*	0.04*	.33	.37
Europe ^a	18	6	2,965	1.35	0.30	0.53								
Africa ^a	9	4	3,303	1.12	0.12	0.22								
Asia	90	7	5,177	1.26	0.23	0.59	1.37	0.32*	0.10	[0.11, 0.52]	0.01	0.05	.13	.44

Note. k_1 = Number of effect sizes; k_2 = Number of samples; N = Total sample size; OR = Mean unweighted odds ratio; Ω = Aggregated, inverse variance-weighted odds ratio; SE = Standard error of log Ω ; 95%CI = 95% confidence interval of log Ω ; τ^2 = Random level 2 or level 3 variance of log Ω ; I^2 = Proportion of total variance in log Ω due to level 2 or level 3 between-study heterogeneity (Cheung, 2013); FSN = Fail safe number of null effects (Rosenberg, 2005); p = Significance level of regression test for funnel plot asymmetry (Peters et al., 2006). Positive log Ω indicate higher prevalence rates in computerized assessments.

^a Estimation problems of multilevel model because of small number of available studies.

* $p < .05$, [†] $p < .10$; [†] Robust FSN > $5k_1 + 10$ (Rosenthal, 1979)

Table 5 (continued)

	Publication bias	
	FSN	<i>p</i>
Overall	11,575 [†]	.28
<i>Type of sensitive behavior</i>		
Substance use	4,024 [†]	.17
Sexuality	142	.45
Delinquency	0	.70
Victimization	0	.58
<i>Geographical region</i>		
United States	5,098 [†]	.86
Europe ^a		
Africa ^a		
Asia	1,352 [†]	.06

Table 6

Moderator Analyses for Sensitive Behaviors in Computerized Assessments

	Model 1		Model 2		Model 3a		Model 3b		Model 4	
	(Survey year)		(Item sensitivity)		(Procedural characteristics)		(Sample characteristics)			
	Ω	γ (SE)	Ω	γ (SE)	Ω	γ (SE)	Ω	γ (SE)	Ω	γ (SE)
Intercept (γ_0)	1.41	0.34* (0.14)	1.53	0.42* (0.15)	1.38	0.32* (0.15)	1.51	0.41* (0.16)	1.29	0.26* (0.12)
Random level 2 variance $\tau^2_{(2)}$		0.03* (0.01)		0.02* (0.01)		0.03* (0.01)		0.02* (0.01)		0.01 (0.00)
Random level 3 variance $\tau^2_{(3)}$		0.03* (0.01)		0.04* (0.02)		0.01* (0.01)		0.02* (0.01)		0.01 (0.00)
1. Survey year: linear (γ_1)		0.07+ (0.04)		0.08+ (0.04)		0.10* (0.04)		0.11* (0.04)		0.05 (0.00)
quadratic (γ_2)		0.00* (0.00)		0.01* (0.00)		0.01* (0.00)		0.01* (0.00)		0.00 (0.01)
Year 1995	1.25		1.32		1.08		1.30		1.39	
Year 2000	1.08		1.12		0.93		0.96		1.14	
Year 2005	1.19		1.26		1.09		1.14		1.16	
2. Item sensitivity (γ_3)				0.02* (0.00)				0.02* (0.00)		0.01* (0.01)
Upper quartile			1.63				1.61		1.35	
Lower quartile			1.43				1.41		1.24	
3. Interviewer presence (γ_4)						0.10 (0.05)		0.08 (0.06)		
Present					1.52		1.62			
Not present					1.25		1.39			
4. Group administration (γ_5)						-0.16* (0.05)		-0.20* (0.06)		
Group					1.18		1.24			
Individual					1.61		1.83			
5. Survey setting (γ_6)						-0.04 (0.04)		-0.02 (0.05)		
Standardized					1.32		1.48			
Unstandardized					1.44		1.53			

Table 6 (continued)

	Model 1		Model 2		Model 3a		Model 3b		Model 4	
	(Survey year)		(Item sensitivity)		(Procedural characteristics)				(Sample characteristics)	
	Ω	γ (<i>SE</i>)	Ω	γ (<i>SE</i>)	Ω	γ (<i>SE</i>)	Ω	γ (<i>SE</i>)	Ω	γ (<i>SE</i>)
		:		:		:		:		
		:		:		:		:		
6. Audio-enhancement (γ_7)										
included						-0.01 (0.05)		-0.03 (0.05)		
not included					1.36		1.47			
					1.40		1.55			
7. Sex of respondents (γ_8)										.00 (.00)
Male										1.34
Female										1.24
8. Age of respondents (γ_9)										.01 (.00)
15 years										1.29
30 years										1.45
$N_{\text{Level 2}} / N_{\text{Level 3}}$	343 / 31		283 / 29		336 / 29		278 / 27		138 / 16	
$R^2_{\text{Level 2}} / R^2_{\text{Level 3}}$.00 / .13		.22 / .00		.00 / .63		.22 / .55		.22 / .46	

Note. γ_0 = Intercept representing the aggregated, true logOR after correcting for moderators; Ω = Predicted true odds ratio; γ = Fixed effects weight; *SE* = Standard error of γ .

* $p < .05$, + $p < .06$.

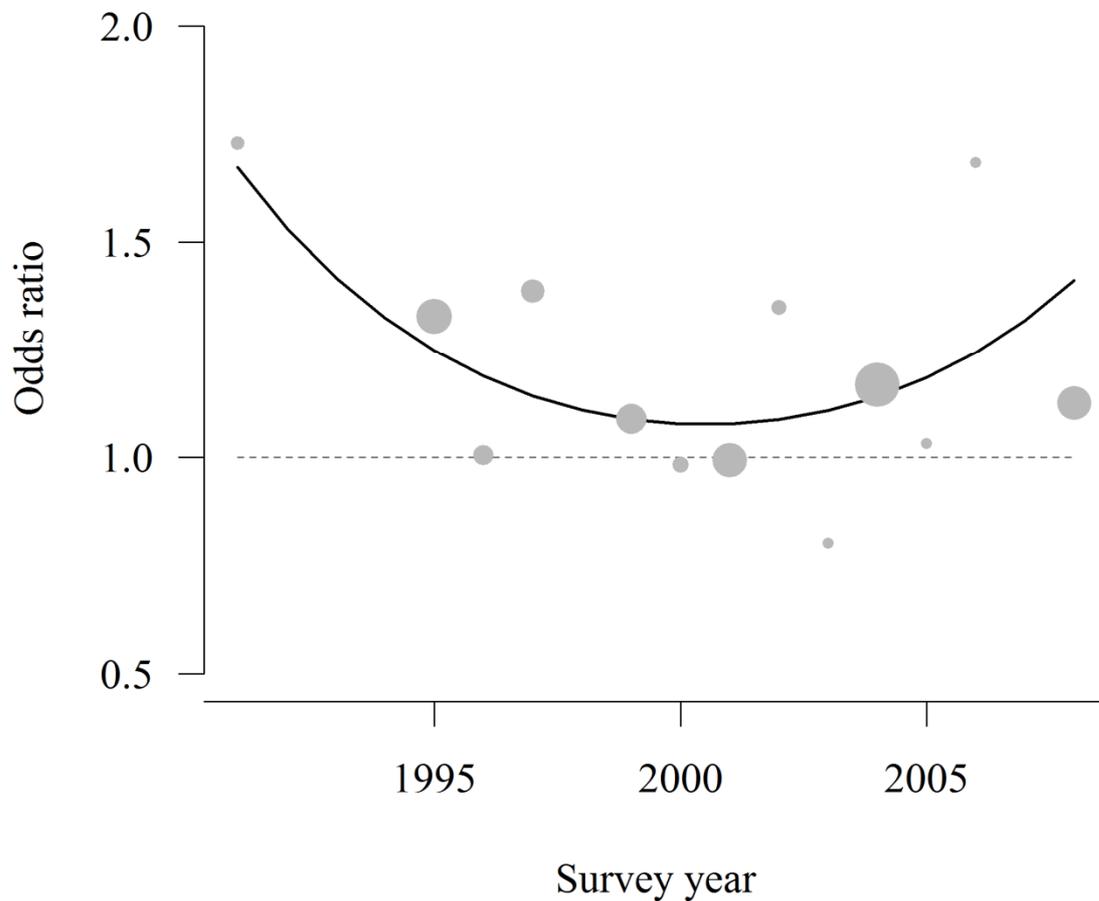


Figure 1. Effect of computerized assessment on self-disclosure across time. Odds ratios greater 1 indicate higher prevalence rates of self-reported sensitive behaviors in computerized than in paper-and-pencil surveys. The solid line represents the model implied change trajectory from regression 1 in Table 6; dots represent the aggregated true effects for the respective year (dot sizes correspond to the number of included effects).

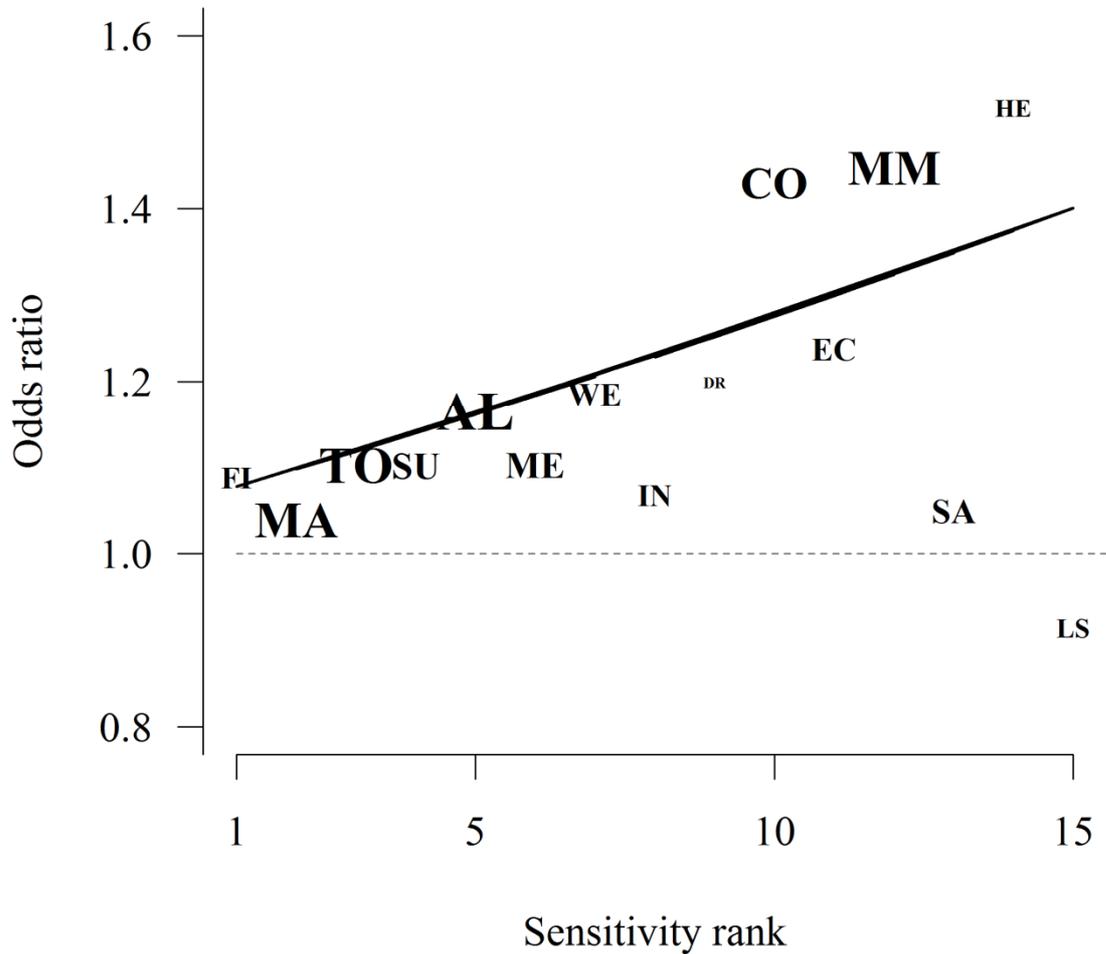


Figure 2. Effect of computerized assessment on self-disclosure by sensitivity of behavior. Odds ratios greater 1 indicate higher prevalence rates of self-reported sensitive behaviors in computerized than in paper-and-pencil surveys. The solid line represents the regression line. Letters indicate the mean effects for different types of sensitive behavior (for abbreviations see Table 2); font sizes correspond to the number of included effects.

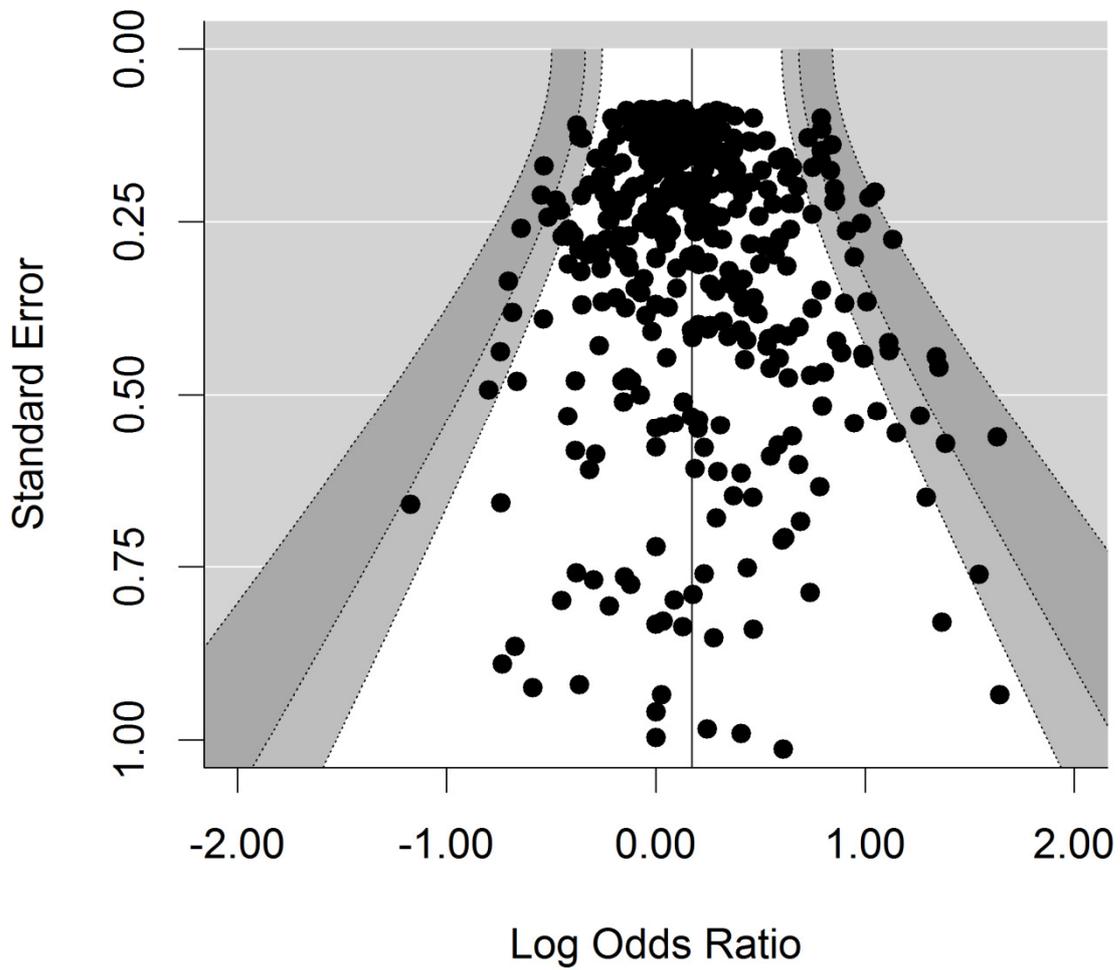


Figure 3. Contour-enhanced funnel plots with 90% (white), 95% (light gray), and 99% (dark gray) confidence intervals around the aggregated true effect (horizontal line).